

STIC-Biotech/ChemLib

31136

From: Hutzell, Paula
Sent: Tuesday, December 12, 2000 5:03 PM
To: STIC-Biotech/ChemLib; Bui, Phuong
Subject: FW: 09/301906 sequence search

P

approved

-----Original Message-----

From: Bui, Phuong
Sent: Tuesday, December 12, 2000 3:15 PM
To: Hutzell, Paula
Subject: 09/301906 sequence search

FD 4/29/99
PTD 4/29/98

Paula,

Please forward the following search to **Barb O'Bryen** as a rush. Thanks. P.

Applic. No. 09/301906
Phuong Bui
AU 1638
CM1, 9A09
305-1996

Please search SEQ ID NO. 4. Please include interference search.

Thank you! Phuong.

Scientific and Technical Information Center

3 11 36

Requester's Full Name: _____ Examiner #: _____ Date: _____
 Art Unit: _____ Phone Number 30 _____ Serial Number: _____
 Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

 Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

STAFF USE ONLY

Searcher: 2003
 Searcher Phone #: _____
 Searcher Location: _____
 Date Searcher Picked Up: 12-18-00
 Date Completed: 12-18-00
 Searcher Prep & Review Time: _____
 Clerical Prep Time: _____
 Online Time: _____

Type of Search

NA Sequence (#) _____
 AA Sequence (#) _____
 Structure (#) _____
 Bibliographic _____
 Litigation _____
 Fulltext _____
 Patent Family _____
 Other _____

Vendors and cost where applicable

STN _____
 Dialog _____
 Questel/Orbit _____
 Dr.Link _____
 Lexis/Nexis _____
 Sequence Systems _____
 WWW/Internet _____
 Other (specify) _____

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 18, 2000, 10:00:33 ; Search time 1016.61 Seconds
(without alignments)
1546.639 Million cell updates/sec

Title: US-09-301-906-4
Perfect score: 360
Sequence: 1 gtcagcggtcagtcagcgc.....agcgcgcttgacgtttggc 360

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues
Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba1:*
- 2: gb_ba2:*
- 3: gb_em:*
- 4: gb_ov:*
- 5: gb_pat:*
- 6: gb_ph:*
- 7: gb_pl1:*
- 8: gb_pl2:*
- 9: gb_pl3:*
- 10: gb_pr2:*
- 11: gb_pr3:*
- 12: em_fun:*
- 13: em_hum1:*
- 14: em_hum2:*
- 15: em_in:*
- 16: em_om:*
- 17: em_or:*
- 18: em_ov:*
- 19: em_pat:*
- 20: em_ph:*
- 21: em_pl:*
- 22: em_ro:*
- 23: em_sts:*
- 24: em_sy:*
- 25: em_un:*
- 26: em_vi:*
- 27: gb_htg1:*
- 28: gb_htg2:*
- 29: gb_in1:*
- 30: gb_in2:*
- 31: em_ba1:*
- 32: em_ba2:*
- 33: em_hum3:*
- 34: em_hum4:*
- 35: gb_pr4:*
- 36: gb_htg3:*
- 37: gb_htg4:*
- 38: gb_htg5:*
- 39: gb_htg6:*
- 40: gb_htg7:*
- 41: em_htg1:*
- 42: em_htg2:*
- 43: em_htg3:*

- 44: em_hum5:*
- 45: gb_pl3:*
- 46: gb_pr5:*
- 47: gb_htg8:*
- 48: gb_htg9:*
- 49: gb_htg10:*
- 50: gb_htg11:*
- 51: gb_htg12:*
- 52: gb_htg13:*
- 53: gb_htg14:*
- 54: gb_in3:*
- 55: gb_htg15:*
- 56: gb_htg16:*
- 57: gb_htg17:*
- 58: em_htg4:*
- 59: em_htg5:*
- 60: em_htg6:*
- 61: em_htg7:*
- 62: em_hum6:*
- 63: gb_htg18:*
- 64: gb_htg19:*
- 65: gb_ba3:*
- 66: em_htg8:*
- 67: em_htg9:*
- 68: em_htg10:*
- 69: em_htg11:*
- 70: em_htg12:*
- 71: em_htg13:*
- 72: em_htg14:*
- 73: em_htg15:*
- 74: em_htg16:*
- 75: em_htg17:*
- 76: em_htg18:*
- 77: em_htg19:*
- 78: em_htg20:*
- 79: em_htg21:*
- 80: em_htg22:*
- 81: em_htg23:*
- 82: gb_pr6:*
- 83: gb_pr7:*
- 84: gb_htg20:*
- 85: gb_htg21:*
- 86: gb_htg22:*
- 87: gb_htg23:*
- 88: gb_ro:*
- 89: gb_sts1:*
- 90: gb_sts2:*
- 91: gb_sy:*
- 92: gb_un:*
- 93: gb_vil:*
- 94: gb_vil2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Match	Length	DB	ID	Description
1	39	10.8	3782	54	DROELAVG		M61748 Drosophila
2	38	10.6	1287	45	OSPER		X66125 O.sativa mr
3	38	10.6	1306	7	AF014467		AF014467 Orvza sat
C	38	10.6	1611	54	TBRPL1		Z54340 T.brucei ge
5	36.8	10.2	677	3	AF219257		AF219257 Bos tauru
6	35.6	9.9	2140	4	XELMAPKK		D13700 Xenopus lae
7	35.2	9.8	384	11	AF219258		AF219258 Homo sapi
C	35	9.7	160013	11	AL133325		AL133325 Human DNA
C	34.2	9.5	753	88	MMU18673		U18673 Mus musculu
C	33.8	9.4	9003	38	AC014530		AC014530 Drosophil
C	33.8	9.4	83188	37	AC013383		AC013383 Homo sapi
C	33.8	9.4	143816	38	AC013598		AC013598 Homo sapi

Db 121 GCAGAAAAAGTTGGAGAGCTTGAGC 146

RESULT 7

AF219258 384 bp mRNA PRI 09-FEB-2000

LOCUS Homo sapiens synoretin mRNA, complete cds.

DEFINITION AF219258

ACCESSION AF219258.1 GI:6942175

VERSION

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 384)

AUTHORS Surguchov,A., Surgucheva,I., Solessio,E. and Baehr,W.

TITLE Synoretin: A new protein belonging to the synuclein family

JOURNAL Mol. Cell. Neurosci. 13 (2), 95-103 (1999)

MEDLINE 99210388

REFERENCE 2 (bases 1 to 384)

AUTHORS Surguchov,A., Surgucheva,I., Baehr,W. and Solessio,E.

TITLE Direct Submission

JOURNAL Submitted (27-DEC-1999) Ophthalmology, Washington University School of Medicine in St. Louis, 660 South Euclid, St. Louis, MO 63108, USA

FEATURES

source Location/Qualifiers

1..384

/organism="Homo sapiens"

/db_xref="taxon:9606"

1..384

/note="member of gamma-subfamily of synuclein; similar to persyn"

/codon_start=1

/product="synoretin"

/protein_id="AAF32343.1"

/db_xref="GI:6942176"

/translation="MDVFKKGFSLAKGVGAVETKOGVTAAEKTKEGVMYVGAKT KEGVQSVTSVAETKQANAVSEAVSVNTVATKIVEENIAVTSGVVHKEALKQ PVPQDEAKAEQVAETKSGSD"

BASE COUNT 104 a 75 c 142 g 63 t

ORIGIN

Query Match 9.8%; Score 35.2; DB 11; Length 384;

Best Local Similarity 50.6%; Pred. No. 10;

Matches 85; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

Qy 185 ctatccgcgtgttaagcaagatgacgagagcccaacgcaagcgtcacattggcgtg 244

Db 32 CCAAGGAGGGTGTGGTGGTCCCTGGGAAGACCAAGCAGGAGTGACAGAGCGAGCTG 91

Qy 245 ttaagccaacagctgtgtggctccacgtaccactctctataaaacaggaagcactggagc 304

Db 92 AGAAGACCAAGGAGGGTGTATGTATGTGGGAGCTAAGACCAAGAGAGGTGTGTGCAGA 151

Qy 305 cagcgcaatcagtcaccacacacagctgttgaggagagagagcgccttga 352

Db 152 GTGTCACTTCAGTGGCTGAGAAGACCAAGAGGAGCAGGCCAACGCCGTGA 199

RESULT 8

AL133325/c

LOCUS AL133325 160013 bp DNA PRI 01-AUG-2000

DEFINITION Human DNA sequence from clone RP5-984P4 on chromosome 20 Contains a gene encoding the homeobox protein NKX2B, a pseudogene similar to steroid dehydrogenase, Glutathione S-transferase (GSTM5) pseudogene, CpG islands, ESTs, STSs and GSSs, complete sequence.

ACCESSION AL133325

VERSION AL133325.20 GI:8217438

KEYWORDS HTG; CpG island; NKX2B.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 160013)

AUTHORS Clark,G.

TITLE Direct Submission

JOURNAL Submitted (01-AUG-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk

COMMENT On Jun 3, 2000 this sequence version replaced gi:8051767. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr20 RP5-984P4 is from the library RPCI-5 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/ VECTOR: pcypac2 IMPORTANT: This sequence is not the entire insert of clone RP5-984P4 It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RP5-984P4 is at 1 in this sequence. The true left end of clone RP1-167022 is at 159914 in this sequence. The true right end of clone RP11-227D2 is at 74869 in this sequence.

FEATURES

source Location/Qualifiers

1..160013

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="20"

/clone="RP5-984P4"

/clone_lib="RPCI-5"

78..611

/note="match: GSS: Em:AQ082697"

87..516

/note="match: GSS: Em:AQ799623"

479..773

/note="AluY repeat: matches 17..311 of consensus"

complement(778..1208)

/note="match: STS: Em:HS184G12T"

1174..1556

/note="MSTB repeat: matches 1..407 of consensus"

1674..1741

/note="34 copies 2 mer ta 82% conserved"

1676..1743

/note="17 copies 4 mer tata 82% conserved"

2054..2343

/note="AluX repeat: matches 1..290 of consensus"

2556..2861

/note="AluJb repeat: matches 1..310 of consensus"

3027..3139

/note="MER33 repeat: matches 217..324 of consensus"

3140..3277

/note="FLAM_C repeat: matches 1..133 of consensus"

3278..3470

/note="MER33 repeat: matches 5..217 of consensus"

3483..3806

/note="MERS8B repeat: matches 1..341 of consensus"

J. Virol. 69 (4), 2501-2507 (1995)
 95191029
 2 (bases 1 to 753)
 Shackleford,G.M.
 Direct Submission
 Submitted (15-DEC-1994) Gregory M. Shackleford, Division
 Hematology-Oncology, Univ. Of Southern California and Childrens
 Hospital Los Angeles, 4650 Sunset Boulevard, Los Angeles, CA
 90027-6016, USA

FEATURES

source	Location/Qualifiers	
	1..753	
	/organism="Mus musculus"	
	/db_xref="taxon:10090"	
	/sex="female"	
	/tissue_type="mammary tumor"	
gene	85..699	
	/gene="Fgf-8"	
CDS	85..699	
	/gene="Fgf-8"	
	/note="alternatively spliced mRNA; isoform Fgf-8a"	
	/codon_start=1	
	/product="fibroblast growth factor-8a"	
	/protein_id="AAA65387.1"	
	/db_xref="GI:619920"	
	[translation="MGSPRSALSCLLHLVLQAHVREQSLVDQLSRRLIRTYQ LYSTSGKHVOVLANKRINAMADGDPFAKLIVETDTFGSRVRGAETGLYICMKNK GKLIASKNGKDKCVTEILVLENNYTALONAKYEGWMAFTRKGRPRKSGKTQHORE VHPMKRLPRGHHTTSESLAFEFNLNYPPTFRSLGSORTWAPEPR"]	
BASE COUNT	152 a	265 c 227 g 109 t
ORIGIN		

```

Query Match          9.5%;   Score 34.2;   DB 88;   Length 753;
Best Local Similarity 48.2%;   Pred. No. 21;
Matches 96; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY  148  cgggtagtcgctatgccccgcctcaacgcgcgcagcctaaccgcgtgttaagaagaat 207
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  207  CCGGATAGGGCGGGCTGACGTGCCTCCGTACCAGGCTCTGTCCTCACAATCGTGCGC 148
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY  208  gagcagaagccccaaacaggcagcgctcacattgggtgtttaagccaacagtgttgcgctc 267
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  147  TTGGAGCAGAGAACCAGCAAGTCAACAGCAGCAGCAGCTCAGCGCGGAGCGGGGCTGCC 88
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY  268  cagctaacattctctaaaaaagggaagcactggcagcgcgaatacgtcccacaacag 327
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  87  CATGGCGCGGGCCCCGGGAGACTGAGAGCCCAGCGGGGTCTACGCGCTCCCATCGGAG 28
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY  328  tcgttgaggaggaagccg 346
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  27  GCCCGGGGGGACAAAGCCG 9
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
AC014530/c

LOCUS       AC014530           9003 bp    DNA             HTG              16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC014530
VERSION     AC014530.1 GI:6436805
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
ORGANISM    Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 9003)
AUTHORS    Adams,M. and Venter,J.C.
TITLE      Direct Submission
JOURNAL    Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT    This sequence was identified as CDM:10212184 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
```

```

* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
    source
        1..9003
        /organism="Drosophila melanogaster"
        /db_xref="taxon:7227"
BASE COUNT      2485 a  1945 c  1940 g  2633 t
ORIGIN

Query Match      9.4%; Score 33.8; DB 38; Length 9003;
Best Local Similarity 49.7%; Pred. NO. 30;
Matches 86; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY  170 tcaagcgcgagcagcctaaaccgctgttaagcaaatgaagcagaagcccaaacgcgaag 229
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  8682 TCAAGCGCTCTGCAGCAGCAGCAGCGCCATGCAACAGCAGCAGCAGCAGCATCAAC 8623
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY  230 cgtcacattggcgtgttaagcccaacagctgttgctgcgcgtaccacttccataaaaac 289
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  8622 AACACAGCAGCAGCAGCAGCAGCCCAAAATGGTGGTTCCTCAATTCACAAACACACAGC 8563
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY  290 aggaagcactggagccagcgaatcagtcaccacacagctcgttggaggagaag 342
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  8562 AGCAGCAACAGGTGCCCGCAGCAGCACTGCAGCAGCAGCAGCAGCAGCAGCAG 8510
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
AC013383
LOCUS      Homo sapiens chromosome 2 clone RP11-396G4 map 2, LOW-PASS SEQUENCE
DEFINITION
ACCESSION  AC013383
VERSION    AC013383.2 GI:9123916
KEYWORDS   HTG; HTGS_PHASE0.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 83188)
AUTHORS    Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE      Homo sapiens chromosome 2, clone RP11-396G4
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 83188)
AUTHORS    Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguski,K., Bouckhvalter,B.,
            Brown,A., Castle,A., Collangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArelano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McDurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
            Direct Submission
            Submitted (09-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            On Jul 13, 2000 this sequence version replaced gi:6289163.
            All repeats were identified using RepeatMasker:
            Smit,A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ----- Project Information

```

Center project name: L3790
Center clone name: 396_G_4

* NOTE: This record contains 87 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 635: contig of 635 bp in length
* 636 735: gap of 100 bp
* 736 1579: contig of 844 bp in length
* 1580 1679: gap of 100 bp
* 1680 2558: contig of 879 bp in length
* 2559 2658: gap of 100 bp
* 2659 3511: contig of 853 bp in length
* 3512 3611: gap of 100 bp
* 3612 4461: contig of 850 bp in length
* 4462 4561: gap of 100 bp
* 4562 5399: contig of 838 bp in length
* 5400 5499: gap of 100 bp
* 5500 6399: contig of 900 bp in length
* 6400 6499: gap of 100 bp
* 6500 7355: contig of 856 bp in length
* 7356 7455: gap of 100 bp
* 7456 8268: contig of 813 bp in length
* 8269 8368: gap of 100 bp
* 8369 9239: contig of 871 bp in length
* 9240 9339: gap of 100 bp
* 9340 10190: contig of 851 bp in length
* 10191 10290: gap of 100 bp
* 10291 11147: contig of 857 bp in length
* 11148 11247: gap of 100 bp
* 11248 12131: contig of 884 bp in length
* 12132 12231: gap of 100 bp
* 12232 13105: contig of 874 bp in length
* 13106 13205: gap of 100 bp
* 13206 14071: contig of 866 bp in length
* 14072 14171: gap of 100 bp
* 14172 15049: contig of 878 bp in length
* 15050 15149: gap of 100 bp
* 15150 16013: contig of 864 bp in length
* 16014 16113: gap of 100 bp
* 16114 16981: contig of 868 bp in length
* 16982 17081: gap of 100 bp
* 17082 17942: contig of 861 bp in length
* 17943 18042: gap of 100 bp
* 18043 18903: contig of 861 bp in length
* 18904 19003: gap of 100 bp
* 19004 19879: contig of 876 bp in length
* 19880 19979: gap of 100 bp
* 19980 20834: contig of 855 bp in length
* 20835 20934: gap of 100 bp
* 20935 21787: contig of 853 bp in length
* 21788 21887: gap of 100 bp
* 21888 22742: contig of 855 bp in length
* 22743 22842: gap of 100 bp
* 22843 23691: contig of 849 bp in length
* 23692 23791: gap of 100 bp
* 23792 24657: contig of 866 bp in length
* 24658 24757: gap of 100 bp
* 24758 25630: contig of 873 bp in length
* 25631 25730: gap of 100 bp
* 25731 26584: contig of 854 bp in length
* 26585 26684: gap of 100 bp
* 26685 27579: contig of 895 bp in length
* 27580 27679: gap of 100 bp
* 27680 28553: contig of 874 bp in length

28554 28653: gap of 100 bp
28654 29505: contig of 852 bp in length
29506 29605: gap of 100 bp
29606 30433: contig of 828 bp in length
30434 30533: gap of 100 bp
30534 31389: contig of 856 bp in length
31390 31489: gap of 100 bp
31490 32354: contig of 865 bp in length
32355 32454: gap of 100 bp
32455 33314: contig of 860 bp in length
33315 33414: gap of 100 bp
33415 34265: contig of 851 bp in length
34266 34365: gap of 100 bp
34366 35234: contig of 869 bp in length
35235 35334: gap of 100 bp
35335 36206: contig of 872 bp in length
36207 36306: gap of 100 bp
36307 37170: contig of 864 bp in length
37171 37270: gap of 100 bp
37271 38143: contig of 873 bp in length
38144 38243: gap of 100 bp
38244 39107: contig of 864 bp in length
39108 39207: gap of 100 bp
39208 40060: contig of 853 bp in length
40061 40160: gap of 100 bp
40161 41019: contig of 859 bp in length
41020 41119: gap of 100 bp
41120 41975: contig of 856 bp in length
41976 42075: gap of 100 bp
42076 42931: contig of 856 bp in length
42932 43031: gap of 100 bp
43032 43864: contig of 833 bp in length
43865 43964: gap of 100 bp
43965 44825: contig of 861 bp in length
44826 44925: gap of 100 bp
44926 45813: contig of 888 bp in length
45814 45913: gap of 100 bp
45914 46779: contig of 866 bp in length
46780 46879: gap of 100 bp
46880 47715: contig of 836 bp in length
47716 47815: gap of 100 bp
47816 48694: contig of 879 bp in length
48695 48794: gap of 100 bp
48795 49621: contig of 827 bp in length
49622 49721: gap of 100 bp
49722 50598: contig of 877 bp in length
50599 50698: gap of 100 bp
50699 51562: contig of 864 bp in length
51563 51662: gap of 100 bp
51663 52540: contig of 878 bp in length
52541 52640: gap of 100 bp
52641 53513: contig of 873 bp in length
53514 53613: gap of 100 bp
53614 54474: contig of 861 bp in length
54475 54574: gap of 100 bp
54575 55434: contig of 860 bp in length
55435 55534: gap of 100 bp
55535 56424: contig of 890 bp in length
56425 56524: gap of 100 bp
56525 57450: contig of 926 bp in length
57451 57550: gap of 100 bp
57551 58410: contig of 860 bp in length
58411 58510: gap of 100 bp
58511 59329: contig of 819 bp in length
59330 59429: gap of 100 bp
59430 60276: contig of 847 bp in length
60277 60376: gap of 100 bp
60377 61245: contig of 869 bp in length
61246 61345: gap of 100 bp
61346 62198: contig of 853 bp in length
62199 62298: gap of 100 bp
62299 63171: contig of 873 bp in length
63172 63271: gap of 100 bp


```
* 29022 29121: gap of 100 bp
* 29122 29998: contig of 877 bp in length
* 29999 30098: gap of 100 bp
* 30099 30974: contig of 876 bp in length
* 30975 31074: gap of 100 bp
* 31075 31958: contig of 884 bp in length
* 31959 32058: gap of 100 bp
* 32059 32934: contig of 876 bp in length
* 32935 33034: gap of 100 bp
* 33035 33898: contig of 864 bp in length
* 33899 33998: gap of 100 bp
* 33999 34825: contig of 827 bp in length
* 34826 34925: gap of 100 bp
* 34926 35779: contig of 854 bp in length
* 35780 35879: gap of 100 bp
* 35880 36743: contig of 864 bp in length
* 36744 36843: gap of 100 bp
* 36844 37702: contig of 859 bp in length
* 37703 37802: gap of 100 bp
* 37803 38687: contig of 885 bp in length
* 38688 38787: gap of 100 bp
* 38788 39672: contig of 885 bp in length
* 39673 39772: gap of 100 bp
* 39773 40678: contig of 906 bp in length
* 40679 40778: gap of 100 bp
* 40779 41650: contig of 872 bp in length
* 41651 41750: gap of 100 bp
* 41751 42622: contig of 872 bp in length
* 42623 42722: gap of 100 bp
* 42723 43563: contig of 841 bp in length
* 43564 43663: gap of 100 bp
* 43664 44531: contig of 868 bp in length
* 44532 44631: gap of 100 bp
* 44632 45490: contig of 859 bp in length
* 45491 45590: gap of 100 bp
* 45591 46469: contig of 879 bp in length
* 46470 46569: gap of 100 bp
* 46570 47423: contig of 854 bp in length
* 47424 47523: gap of 100 bp
* 47524 48469: contig of 946 bp in length
* 48470 48569: gap of 100 bp
* 48570 49398: contig of 829 bp in length
* 49399 49498: gap of 100 bp
* 49499 50355: contig of 857 bp in length
* 50356 50455: gap of 100 bp
* 50456 51311: contig of 856 bp in length
* 51312 51411: gap of 100 bp
* 51412 52275: contig of 864 bp in length
* 52276 52375: gap of 100 bp
* 52376 53267: contig of 892 bp in length
* 53268 53367: gap of 100 bp
* 53368 54211: contig of 844 bp in length
* 54212 54311: gap of 100 bp
* 54312 55200: contig of 889 bp in length
* 55201 55300: gap of 100 bp
* 55301 56160: contig of 860 bp in length
* 56161 56260: gap of 100 bp
* 56261 57119: contig of 859 bp in length
* 57120 57219: gap of 100 bp
* 57220 58110: contig of 891 bp in length
* 58111 58210: gap of 100 bp
* 58211 59096: contig of 886 bp in length
* 59097 59196: gap of 100 bp
* 59197 60038: contig of 842 bp in length
* 60039 60138: gap of 100 bp
* 60139 61010: contig of 872 bp in length
* 61011 61110: gap of 100 bp
* 61111 61960: contig of 850 bp in length
* 61961 62060: gap of 100 bp
* 62061 62929: contig of 869 bp in length
* 62930 63029: gap of 100 bp
* 63030 63902: contig of 873 bp in length
* 63903 64002: gap of 100 bp
```

```
* 64003 64863: contig of 861 bp in length
* 64864 64963: gap of 100 bp
* 64964 65826: contig of 863 bp in length
* 65827 65926: gap of 100 bp
* 65927 66798: contig of 872 bp in length
* 66799 66898: gap of 100 bp
* 66899 67769: contig of 871 bp in length
* 67770 67869: gap of 100 bp
* 67870 68729: contig of 860 bp in length
* 68730 68829: gap of 100 bp
* 68830 69697: contig of 868 bp in length
* 69698 69797: gap of 100 bp
* 69798 70653: contig of 856 bp in length

Query Match 9.4%; Score 33.8; DB 38; Length 143816;
Best Local Similarity 50.3%; Pred. No. 33;
Matches 83; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 114 tgtttcattgacttctgtacacaccacccgggtagtcgctatgcccccgtcaa 173
|| || || || || || || || || || || || || || || || || || ||
Db 6145 TGCCTTTAGGCCATTATCTGTAATAATACACAGTCTCTTTGGTTGCTAGTAAAA 6086

QY 174 gcgcgagcagcctaaacccgctgttaagcaagatgacagagcccaagcgcgcgc 233
|| || || || || || || || || || || || || || || || || || ||
Db 6085 CAATGAGGACACACACCTTAAGAAATGTAACGAGACCCAAAGACAAACGGGAATGTC 6026

QY 234 acattgggctgttaagcccaacgctgttgccgctccacgtaccact 278
|| || || || || || || || || || || || || || || || || || ||
Db 6025 TCTCTTTGCCAGACACCTTCTCTGTACACCTGCACATAGCCCT 5981

RESULT 13
AC011243
LOCUS
DEFINITION
AC011243 Homo sapiens chromosome 2 clone RP11-491M19, WORKING DRAFT
AC011243 HTG
AC011243.5 GI:9795966
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Waterston,R.H.
1 (bases 1 to 174101)
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 174101)
Waterston,R.H.
Direct Submission
Submitted (04-OCT-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Aug 11, 2000 this sequence version replaced gi:8568876.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0491M19
----- Summary Statistics -----
Sequencing vector: M13; 68%
Chemistry: Dye-terminator Big Dye; 60% of reads
Chemistry: Dye-terminator Big Dye; 40% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 165435 bases at least Q40
Consensus quality: 168680 bases at least Q30
Consensus quality: 170448 bases at least Q20
Insert size: 175000; agarose-fp
Insert size: 173001; sum-of-contigs
Quality coverage: 4.47 in Q20 bases; agarose-fp
```

Quality coverage: 4.66 in Q20 bases; sum-of-contigs

***** NOTE: This is a 'working draft' sequence. It currently consists of 9 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 3337: contig of 3337 bp in length
 3338 3437: gap of unknown length
 3438 10367: contig of 6930 bp in length
 10368 10467: gap of unknown length
 10468 22403: contig of 11936 bp in length
 22404 22503: gap of unknown length
 22504 31981: contig of 9478 bp in length
 31982 32081: gap of unknown length
 32082 44508: contig of 12427 bp in length
 44509 44608: gap of unknown length
 44609 71086: contig of 26478 bp in length
 71087 71186: gap of unknown length
 71187 99473: contig of 28287 bp in length
 99474 99573: gap of unknown length
 99574 134831: contig of 35258 bp in length
 134832 134931: gap of unknown length
 134932 174101: contig of 39170 bp in length.

FEATURES

source

1..174101
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /clone="RP11-491M19"

misc_feature

1..3337
 /note="assembly_name:Contig21
 clone_end:SP6
 vector_side:right"

misc_feature

3438..10367
 /note="assembly_name:Contig22"

misc_feature

10468..22403
 /note="assembly_name:Contig23"

misc_feature

22504..31981
 /note="assembly_name:Contig24"

misc_feature

32082..44508
 /note="assembly_name:Contig25"

misc_feature

44609..71086
 /note="assembly_name:Contig26"

misc_feature

71187..99473
 /note="assembly_name:Contig27"

misc_feature

99574..134831
 /note="assembly_name:Contig28"

misc_feature

134932..174101
 /note="assembly_name:Contig29"

BASE COUNT 51725 a 35565 c 34703 g 51300 t 808 others
 ORIGIN

Query Match

Best Local Similarity 9.4%; Score 33.8; DB 36; Length 174101;

Matches 83; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

Qy 114 tttcttaagtacttctgtctacactaccacgggtagtcgctatgccccctcaaa 173

Db 82749 TCGTTTATAGCCATATCTATTAATAATACACAGAGTCTCTTTGGTGTAGTGA 82808

Qy 174 gcgcagagccttaaccggtctttaagcaagatgagcagaagcccaaacggaacgctc 233

Db 82809 CAATGAGGAACACACACCTTATAGAATGTACAGAGAGCAACAAACGGAATGTC 82868

Qy 234 acattggggctgttaagcccaacagctgttggcggtccacgtaccact 278

Db 82869 TCTCTTTGCCAGACACACCTTCTATCTGTACACCTGCACATAGCCCT 82913

RESULT 14

AC018877

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

COMMENT

AC018877 195347 bp DNA HTG 07-AUG-2000
 Homo sapiens chromosome 2 clone RP11-295K1, WORKING DRAFT SEQUENCE,
 27 unordered pieces.
 AC018877
 AC018877.2 GI:9719807
 HTG; HTGS_PHASE1; HTGS_DRAFT.
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 195347)
 Waterston,R.H.
 The sequence of Homo sapiens clone
 Unpublished
 2 (bases 1 to 195347)
 Waterston,R.H.
 Direct Submission
 Submitted (21-DEC-1999) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 On Aug 7, 2000 this sequence version replaced gi:6624094.

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC

Web site: http://genome.wustl.edu/gsc/index.shtml

----- Project Information -----
 Center project name: HNH0295K01

----- Summary Statistics -----
 Sequencing vector: M13; 63%

Sequencing vector: plasmid; 37%

Chemistry: Dye-primer ET; 63% of reads

Chemistry: Dye-terminator Big Dye; 37% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 176143 bases at least Q40

Consensus quality: 182219 bases at least Q30

Consensus quality: 185619 bases at least Q20

Insert size: 194000; agarose-fp

Insert size: 192747; sum-of-contigs

Quality coverage: 3.36 in Q20 bases; agarose-fp

Quality coverage: 3.45 in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently

consists of 27 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

1 1250: contig of 1250 bp in length

1251 1350: gap of unknown length

1351 2427: contig of 1077 bp in length

2428 2527: gap of unknown length

2528 4162: contig of 1635 bp in length

4163 4262: gap of unknown length

4263 6032: contig of 1770 bp in length

6033 6132: gap of unknown length

6133 9005: contig of 2873 bp in length

9006 11520: contig of 2415 bp in length

11521 11620: gap of unknown length

11621 13993: contig of 2373 bp in length

13994 14093: gap of unknown length

14094 16557: contig of 2464 bp in length

16558 16657: gap of unknown length

16658 19112: contig of 2455 bp in length

19113 19212: gap of unknown length

19213 21091: contig of 1879 bp in length

21092 21191: gap of unknown length

```
* 21192 24344: contig of 3153 bp in length
* 24345 24444: gap of unknown length
* 24445 29678: contig of 5234 bp in length
* 29679 33086: gap of unknown length
* 33087 33186: contig of 3308 bp in length
* 33187 37004: contig of 3818 bp in length
* 37005 37104: gap of unknown length
* 37105 41551: contig of 4447 bp in length
* 41552 41651: gap of unknown length
* 41652 47274: contig of 5623 bp in length
* 47275 52187: gap of unknown length
* 52188 52287: gap of 4813 bp in length
* 52288 59023: contig of 6736 bp in length
* 59024 59123: gap of unknown length
* 59124 64299: contig of 5176 bp in length
* 64300 64399: gap of unknown length
* 64400 71405: contig of 6906 bp in length
* 71406 71306: gap of unknown length
* 71406 77968: contig of 6563 bp in length
* 77969 78068: gap of unknown length
* 78069 89603: contig of 11534 bp in length
* 89603 89702: gap of unknown length
* 89703 102691: contig of 12989 bp in length
* 102692 102791: gap of unknown length
* 102792 117322: contig of 14531 bp in length
* 117323 117422: gap of unknown length
* 117423 139728: contig of 22306 bp in length
* 139729 139828: gap of unknown length
* 139829 163216: contig of 23388 bp in length
* 163217 163316: gap of unknown length
* 163317 195347: contig of 32031 bp in length.
FEATURES
    Location/Qualifiers
        1..195347
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="2"
        /clone="RP11-295K1"
        1..1250
        /note="assembly_name:Contig8"
        1351..2427
        /note="assembly_name:Contig11"
        2528..4162
        /note="assembly_name:Contig13"
        4263..6032
        /note="assembly_name:Contig14"
        6133..9005
        /note="assembly_name:Contig15"
        9106..11520
        /note="assembly_name:Contig16"
        11621..13993
        /note="assembly_name:Contig17"
        14094..16357
        /note="assembly_name:Contig18"
        16658..19112
        /note="assembly_name:Contig19"
        19213..21091
        /note="assembly_name:Contig20"
        21192..24344
        /note="assembly_name:Contig21"
        24445..29678
        /note="assembly_name:Contig22"
        29779..33086
        /note="assembly_name:Contig23"
        clone_end:77
        vector_side:right
        33187..37004
        /note="assembly_name:Contig24"
        37105..41551
        /note="assembly_name:Contig25"
        41652..47274
        /note="assembly_name:Contig26"
```

```
misc_feature 47375..52187
              /note="assembly_name:Contig27"
misc_feature 52288..59023
              /note="assembly_name:Contig28"
misc_feature 59124..64299
              /note="assembly_name:Contig29"
misc_feature 64400..71305
              /note="assembly_name:Contig30"
misc_feature 71406..77968
              /note="assembly_name:Contig31"
              clone_end:SP6
              vector_side:right
misc_feature 78069..89602
              /note="assembly_name:Contig32"
misc_feature 89703..102691
              /note="assembly_name:Contig33"
misc_feature 102792..117322
              /note="assembly_name:Contig34"
misc_feature 117423..139728
              /note="assembly_name:Contig35"
misc_feature 139829..163216
              /note="assembly_name:Contig36"
misc_feature 163317..195347
              /note="assembly_name:Contig37"
BASE COUNT 59640 a 38362 c 38114 g 56602 t 2629 others
ORIGIN
Query Match 9.4%; Score 33.8; DB 40; Length 195347;
Best Local Similarity 50.3%; Pred. No. 33;
Matches 83; Conservative 0; Mismatches 82; Indels 0; Gaps 0;
QY 114 tgtttctaagcattacttctgtacactaccacgggttagtcgctatgccccgcgtaaa 173
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159192 TGCCTTTAGCCCATATCTCTAAATAATCACAAGTCTCTTTGGTTTGTAGTGAATA 159251
QY 174 gcgcgagcgcctaaaccggcgtgttaagcaagatgacagagcccaacggcgaagcgtc 233
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159252 CAATGAGGACACACACCTTAAGAAATGTAACGAGACCAAGACAAACCGGAATGTC 159311
QY 234 caattggcgtgttaagcccaacagcgtgttgcgctccacgtaccact 278
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159312 TCCTTTGCCAGACACACCTTCATCTGTACACCTGCACATAGCCCT 159356
RESULT 15
CEY40B1B
LOCUS
DEFINITION Caenorhabditis elegans cosmid Y40B1B, complete sequence.
ACCESSION AL032636 AL009206
VERSION AL032636.1 GI:3810697
KEYWORDS HTG.
SOURCE
ORGANISM
            Caenorhabditis elegans.
            Caenorhabditis elegans.
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE
1 (bases 1 to 29313)
AUTHORS
TITLE
JOURNAL
MEDLINE
REMARK
            The C.elegans Sequencing Consortium.
            Erratum: [[published errata appear in Science 1999 Jan
            1;283(5398):35 and 1999 Mar 26;283(5410):2103 and 1999 Sep
            3;285(5433):1493]]
REFERENCE
2 (bases 1 to 29313)
AUTHORS
TITLE
JOURNAL
```

COMMENT

Coding sequences below are predicted from computer analysis, using predictions from Genefinder (P. Green, U. Washington), and other available information.

For a graphical representation of this sequence and its analysis see:-

<http://webace.sanger.ac.uk/cgi-bin/display?db=wormaceclass=Sequence&object=y40B1B>
 bin/display?db=wormaceclass=Sequence&object=y40B1B
 sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.
 IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

IMPORTANT: This sequence is not the entire insert of clone Y40B1B. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

The true right end of clone W05H12 is at 102 in this sequence. The true right end of clone Y40B1 is at 29313 in this sequence. The start of this sequence (1..102) overlaps with the end of sequence Z82072.

The end of this sequence (29204..29313) overlaps with the start of sequence AL132847.

FEATURES

Location/Qualifiers

1..29313
 /organism="Caenorhabditis elegans"

/db_xref="taxon:6239"

/chromosome="I"

/clone="Y40B1B"

complement(join(2625..3579,3640..3716))

/gene="Y40B1B.3"

complement(join(2625..3579,3640..3716))

/gene="Y40B1B.3"

/codon_start=1

/protein_id="CAA21605.1"

/db_xref="GI:3880909"

/db_xref="SPTREMBL:Q9XWP5"

/translation="MSTFILLPQKSLQNVLRIMQPLELYFSEVSKTCAHTVDLN LKASRLVLKNGECAMIEAVFKDGNVTGFEISTESADCSEVMYPLVGYEDGSY DSHADFTIKVTLIGLAQVIAHFOIFHKPKISQLYHENVYICDQMLEIKTIDFKIE FWITFDYFLITNREVDFOIHLRLVRLVILCRFVDSHLPNVAASNGT QFKCSTVPKISINMELRHWTGNSRPLEICIRISTEAEAPSEESTYEEVLKGTHYO TEPTDSVKTFPDLNLYHNNLCFNEHLKQTGTGIRVCYKKAANNVVKLPETPFESE VLFQMSLIV"

complement(join(5128..5217,6728..6929,6979..7058))

/gene="Y40B1B.7"

complement(join(5128..5217,6728..6929,6979..7058))

/gene="Y40B1B.7"

/note="cDNA EST yk546c3.3 comes from this gene"

cDNA EST yk372d6.3 comes from this gene

cDNA EST yk372d6.5 comes from this gene

cDNA EST yk546c3.5 comes from this gene

cDNA EST yk562g2.5 comes from this gene"

/codon_start=1

/protein_id="CAA21606.1"

/db_xref="GI:3880910"

/db_xref="SPTREMBL:Q9XWP4"

/translation="MSTGANLLVMNDCKSNRWMTKOEKKHSEIKVKYTLKSTWDKK MELAKKDMVKRVODNIREKQVOERQEKKEKKEKRLNERKAEIVQKTKIKHL KTKYKRLRSLQMRDTQVTK"

join(7514..7595,7679..7823,9391..9684,9733..9790,

9847..9995,10578..10908,12393..12628,12679..12798,

12885..12972)

/gene="Y40B1B.8"

join(7514..7595,7679..7823,9391..9684,9733..9790,

9847..9995,10578..10908,12393..12628,12679..12798,

12885..12972)

/gene="Y40B1B.8"

/note="cDNA EST yk597b12.3 comes from this gene"

cDNA EST yk235e7.3 comes from this gene

cDNA EST yk235e7.5 comes from this gene
 cDNA EST yk382c6.5 comes from this gene
 cDNA EST yk597b12.5 comes from this gene"

/codon_start=1

/protein_id="CAA21607.2"

/db_xref="GI:5824762"

/db_xref="SPTREMBL:Q9U2J0"

/translation="MPTQFIRNRSAFPOQTNNDDIDFQYDNFSSSHQLPSTAGYS EHSNHFQNHSPPEKATDPLAGALLGSDVITKLSLHPCGVLRRCQVHOFAGS LHLTVTLIPVCNSVAKOGIOTFWKGAIGSSVLMGLTNVTEVLGMLGRLTFVYN GSAEKYKHIVLKGVTFVTMTPEYISSTFETVRESGIGGEDNKILDVLVKGVDRMY FLTSGRDSRRRSIIHLAIPTAGFQWGHYMLQATLYXOFERMAKRYVMKSPAEKTTY HFLQFMAQTSSMVLTDLILPFETVIRHRTIQGTIRLIDNMDTGLSAYSMVKISG YDCVQVLETFEGWALYAGVAGLVLEYSLHOGQLQVLVACFDRGSELLKKAQTQGHV VQSSQITPPIISAKSFSNPGINSFSPSAGALLSPMTPRKHSHTPPKDDPTFTFGETV SQIGSPYQMDPDPNVGSPPNRTRATHDSTQSLPPLQLNIRPSSGGPPFSG"

join(14099..14519,15347..15429,16169..16303)

/gene="Y40B1B.5"

join(14099..14519,15347..15429,16169..16303)

/gene="Y40B1B.5"

/note="predicted using Genefinder"

cDNA EST yk64e5.5 comes from this gene

cDNA EST yk464a3.5 comes from this gene

cDNA EST yk414e6.5 comes from this gene

cDNA EST yk490e10.5 comes from this gene

cDNA EST yk90f3.5 comes from this gene

cDNA EST yk335g9.5 comes from this gene

cDNA EST yk325b12.5 comes from this gene

cDNA EST yk589e10.5 comes from this gene

cDNA EST yk590a9.5 comes from this gene

cDNA EST yk625g8.5 comes from this gene

cDNA EST yk625g9.5 comes from this gene

cDNA EST yk636a11.5 comes from this gene

cDNA EST yk673f12.5 comes from this gene

cDNA EST yk677d6.5 comes from this gene"

/codon_start=1

/protein_id="CAA21603.1"

/db_xref="GI:3880907"

/db_xref="SPTREMBL:Q9XWP7"

/translation="MSDWEDDFEPEVSTFKRAPEPVKIVDAPPPKAAAPAAK TUKAAPTFAWESLGRLETSAREAIQKNLDLALRDFGDDSAEAIKATYANIAS KADFYRWGRVGFLASCACIYGMIKLLTSDTMDTPAEILKMTVTLQIQISAAK KTAETKATKATTAAPAAANKNAKATLVKTGNSDMYDDYQDDQDDYDY"

join(16662..16740,17501..17730,19395..19700,20983..21625,

22437..22635,23692..23868,25405..25782,28224..28524)

/gene="Y40B1B.6"

join(16662..16740,17501..17730,19395..19700,20983..21625,

22437..22635,23692..23868,25405..25782,28224..28524)

/gene="Y40B1B.6"

/note="predicted using Genefinder"

cDNA EST yk122a5.5 comes from this gene

cDNA EST yk141d9.5 comes from this gene

cDNA EST yk290b12.5 comes from this gene

cDNA EST yk251c1.5 comes from this gene

cDNA EST yk174e9.5 comes from this gene

cDNA EST yk669h5.3 comes from this gene

cDNA EST yk379a6.5 comes from this gene

cDNA EST yk225h10.5 comes from this gene

cDNA EST yk533d5.5 comes from this gene

cDNA EST yk584a9.5 comes from this gene

cDNA EST yk606g3.5 comes from this gene"

/codon_start=1

/protein_id="CAA21604.1"

/db_xref="GI:3880908"

/db_xref="SPTREMBL:Q9XWP6"

/translation="MSSDTGSEYLDDEIRGDELGPSIDONALAAASAARLPDRPTD HELAPFPELWEHTAVERVYRSTKLSRFLVDRRSVIVIGAGAAGISAATOLESPGDV QNIVAYLSRHGLINFGYRVYRSTKLSRFLVDRRSVIVIGAGAAGISAATOLESPGDV IYVLRNCGIGRIHSPKSKSGEIMETGGDTLRLKIESPMTALLHQNFEHGVDFDTS VVEGRPLNEEKIHLFDHYKSAHGALNYOAHOCERHDDOGSFSISROQVLENLSMCE RGTLIKYNFKCSLETVARAREHHFNQKRLMTALMAENQLKMEEGLEQDDPVLRL RSLKRDIAATSLKFEFVADAFETADNHWORLNHNPQAKOYMHFSGSEFATFNFMLGPEE YLVGAQLEKVKQFSCDSMQNKENGVAARLTGEGIAELLTQLSEKRLKLDIRLKHRLDIDY

```
SGFEHVLKLVORENDIEMKAAAFVVSTLPIGVLRKKTIIADERAPTFTPSLPDKKVEA
IRNICGSVNCILEDFRVFTANGRFNOFVTVSPNIRKIRGSMNIWSSVPGSKVLCTY
IVGEAMLELPDDVLIQAFNINLQAFGNCPRAPISAHITRWHDELAFAFGSGAFMSL
RTETTSFDVMEPLKSDGMSRVYFAGEHTCSSYSTTQGWMMGARAADISNDHIG
IGFVDISGTRGDRGDERELLIEDVDIGKIEKDENEAVADIPNAPNAPNAQKPEEP
KIAEFIELVAEAEKAEKAEVLEPLVPVE"
Join(Z82072.1:6585..6641,282072.1:7142..7285,
282072.1:7711..7866,282072.1:7985..8185,
282072.1:8682..8861,27..179,229..317,565..874,1630..1743)
/gene="W05H12.1"
CDS
Join(Z82072.1:6585..6641,282072.1:7142..7285,
282072.1:7711..7866,282072.1:7985..8185,
282072.1:8682..8861,27..179,229..317,565..874,1630..1743)
/gene="W05H12.1"
/note="predicted using Genefinder
cDNA EST yk82c3.5 comes from this gene
cDNA EST yk282e6.5 comes from this gene
cDNA EST yk306e5.5 comes from this gene
cDNA EST yk311b10.5 comes from this gene
cDNA EST yk435a2.5 comes from this gene
cDNA EST yk435h2.5 comes from this gene
cDNA EST yk379e4.5 comes from this gene
cDNA EST yk134e7.3 comes from this gene
cDNA EST yk134e7.5 comes from this gene
cDNA EST yk382g1.5 comes from this gene
cDNA EST yk546a9.3 comes from this gene
cDNA EST yk546a9.5 comes from this gene
cDNA EST yk82c3.3 comes from this gene"
Query Match          9.3%; Score 33.4; DB 30; Length 29313;
Best Local Similarity 49.7%; Pred. No. 41;
Matches 85; Conservative 0; Mismatches 86; Indels 0; Gaps 0;
Qy 177 cgagcagcctaaccgctgttaagcaagatgagcagagcccaacgcaagcgtcaca 236
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 616 CGAGGAGGATGAGAGGAAATCTGAGGAAAAAGAGGAAAAAGTTAAGGGAAAAATCAAAACCA 675
Qy 237 ttgggctgttaagccaacagctgttggcgctccacgtaccacttctctaaaaaacaggaagc 296
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 676 GAAAAACAGTAAAAAGAACGAATTCGAGCAGACAGAACGACCAAACTGTAGAAAAAGAGGCA 735
Qy 297 actggagcagcgaatcagtcctcccaacagtcgttgaggaggaagccgc 347
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 736 AGAGAGGCCGAGGAGAGGCCGCTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 786
```

Search completed: December 18, 2000, 15:00:07
Job time: 17974 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 18, 2000, 10:00:33 ; Search time 25.1 Seconds
(without alignments)
2169.194 Million cell updates/sec

Title: US-09-301-906-4

Perfect score: 360

Sequence: 1 gtcagcggctcagtcagcgc.....aggccgcttgagcttggtgc 360

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 262060 seqs, 75620496 residues

Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA:*

1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*

2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*

3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*

4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*

5: /cgn2_6/ptodata/2/ina/6_COMB.seq:*

6: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*

7: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	29.4	8.2	11748	2	US-08-611-107-30
C 2	29	8.1	1560	6	PCT-US94-09789-1
C 3	29	8.1	2158	2	US-08-698-551-1
C 4	29	8.1	2158	3	US-08-602-228-1
C 5	29	8.1	2158	3	US-08-649-341A-1
C 6	29	8.1	2158	3	US-08-494-440B-1
C 7	29	8.1	2158	3	US-08-533-901B-1
C 8	29	8.1	2158	3	US-08-839-032A-1
C 9	29	8.1	2158	4	US-08-839-032A-1
C 10	29	8.1	2158	6	PCT-US95-12724-1
C 11	29	8.1	2735	2	US-08-698-551-11
C 12	29	8.1	2735	3	US-08-602-228-11
C 13	29	8.1	2735	3	US-08-494-440B-11
C 14	29	8.1	2735	3	US-08-533-901B-11
C 15	29	8.1	2735	4	US-08-839-032A-11
C 16	29	8.1	2735	4	US-08-839-032A-11
C 17	29	8.1	2735	6	PCT-US95-12724-11
C 18	29	8.1	6002	2	US-08-698-551-15
C 19	29	8.1	6002	3	US-08-602-228-15
C 20	29	8.1	6002	3	US-08-839-032A-15
C 21	29	8.1	7360	2	US-08-611-107-9
C 22	29	8.1	7360	4	US-08-422-560A-9
C 23	28.2	7.8	545	1	US-07-915-966C-1
C 24	28.2	7.8	545	4	US-08-771-182-1
C 25	28.2	7.8	545	5	US-08-853-194-1
C 26	28	7.8	435	2	US-08-694-579-2

C 27	28	7.8	435	3	US-08-948-155-2	Sequence 2, Appli
28	28	7.8	1689	5	US-08-335-844A-6	Sequence 6, Appli
29	28	7.8	3084	5	US-08-335-844A-19	Sequence 19, Appli
30	27.6	7.7	2012	3	US-08-484-200-3	Sequence 3, Appli
31	27.2	7.6	2223	1	US-08-317-522A-8	Sequence 8, Appli
32	27.2	7.6	2223	1	US-08-439-818A-8	Sequence 8, Appli
33	27.2	7.6	2223	3	US-08-751-965-8	Sequence 8, Appli
34	27.2	7.6	2223	3	US-08-738-975-8	Sequence 8, Appli
35	27.2	7.6	2223	4	US-08-728-626-8	Sequence 8, Appli
36	27.2	7.6	2223	5	US-08-808-599A-8	Sequence 8, Appli
37	26.8	7.4	1524	7	5512669-1	Patent No. 5512669
38	26.8	7.4	1525	7	529279-1	Patent No. 529279
39	26.6	7.4	1045	5	US-08-859-167-3	Sequence 3, Appli
40	26.6	7.4	1045	5	US-09-109-273-3	Sequence 3, Appli
C 41	26.4	7.3	1604	2	US-08-694-579-1	Sequence 1, Appli
C 42	26.4	7.3	1604	3	US-08-948-155-1	Sequence 1, Appli
43	26.2	7.3	1031	1	US-08-181-271A-106	Sequence 106, App
44	26.2	7.3	1031	1	US-08-449-315-106	Sequence 106, App
45	26.2	7.3	1031	1	US-08-444-803-106	Sequence 106, App

ALIGNMENTS

RESULT 1

US-08-611-107-30/c

Sequence 30. Application US/08611107

Patent No. 5801233

GENERAL INFORMATION:

APPLICANT: Haselkorn, Robert

APPLICANT: Gornicki, Piotr

TITLE OF INVENTION: NUCLEIC ACID COMPOSITIONS ENCODING

TITLE OF INVENTION: ACETYL-CoA CARBOXYLASE AND USES

TITLE OF INVENTION: THEREFOR

NUMBER OF SEQUENCES: 40

CORRESPONDENCE ADDRESS:

ADDRESSEE: Arnold, White & Durkee

STREET: P.O. Box 4433

CITY: Houston

STATE: Texas

COUNTRY: United States of America

ZIP: 77210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/611,107

FILING DATE: Concurrently Herewith

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US SN 07/956,700

FILING DATE: 02-OCT-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US PCT/US93/09340

FILING DATE: 30-SEP-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US SN 08/422,560

FILING DATE: 14-APR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Kitchell, Barbara S.

REGISTRATION NUMBER: 33,928

REFERENCE/DOCKET NUMBER: ARCD:221

TELECOMMUNICATION INFORMATION:

TELEPHONE: (512) 418-3000

TELEFAX: (512) 474-7577

INFORMATION FOR SEQ ID NO: 30:

SEQUENCE CHARACTERISTICS:

LENGTH: 11748 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear


```
Query Match      8.1%; Score 29; DB 3; Length 2158;
Best Local Similarity 53.0%; Pred. NO. 1.8;
Matches 62; Conservative 0; Mismatches 55; Indels 0; Gaps
```

5' TCGGGATGTCATTCCTTATTTACCTTCATCAGCAGCATGTAGGAGATGAGGTTGT

RESULT 8

S-08-839-032A-1/C
Sequence 1, Application US/08839032A
Patent No. 5891675
GENERAL INFORMATION.

APPLICANT: Lin, Lih-Ling
 APPLICANT: Chen, Jennifer H.
 APPLICANT: Schlevella, Andrea
 APPLICANT: Crabtree, James

NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:

ADDRESS: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version 1.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/839,032A

REGISTRATION NUMBER: 347/24
REFERENCE/DOCKET NUMBER: G1523DDIV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2158 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 2..1231
S-08-839-032A-1

```
Query Match      8.1%; Score 29; DB 3; Length 2158;
Best Local Similarity 53.0%; Pred. No. 1.8;
Matches 62: Conservative 0; Mismatches 55; Indels
```

[illegible]

RESULT 9

```

US-08-839-031A-1/c
Sequence 1, Application US/08839031A
Patent No. 5948638
GENERAL INFORMATION:
APPLICANT: Lin, Lih-Ling
APPLICANT: Chen, Jennifer H.
APPLICANT: Schievella, Andrea
APPLICANT: Graham, James
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
TITLE OF INVENTION: PROTEINS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/839,031A

```

CLASSIFICATION: 433
ATTORNEY/AGENT INFORMATION:
NAME: Sprunger, Suzanne A.,
REGISTRATION NUMBER: 41,323
REFERENCE/DOCKET NUMBER: G15232BDIV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8284
TELEFAX: (617) 876-5951
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2158 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 2..1231
US-08-839-031A-1

Query Match 8.1%; Score 29; DB 4; Length 2158;
Best Local Similarity 53.0%; Pred. No. 1.8;
Matches 62; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY	195	tgtttaagcaaatgagcagaagcccaaacgcaagcgctcacattggctgtttaagccaac	254
Db	530	TGATTGCTGGCTGTACACAAAGCCCAATGTGCACATTTCCCAATTTAGGCGCCCTCACCTTCT	471
QY	255	agctgttggcgtccacgtaccacttctcaaaaacaggaaagcactggagccacagcga	311
Db	470	TGCGGATGTCATTCTTATTTACCTTCATCAGCAGCATGTAGAGAGTAGAGTTGTGCA	414

RESULT 10
PCT-US95-12724-1/c

```

1 Sequence 1, Application PC/TUS9512724
2
3 GENERAL INFORMATION:
4
5 APPLICANT: Lin, Lih-Ling
6
7 APPLICANT: Chen, Jennifer H.
8
9 APPLICANT: Schievella, Andrea
10
11 APPLICANT: Graham, James
12
13 TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
14
15 TITLE OF INVENTION: PROTEINS AND INHIBITORS OF LIGAND BINDING
16
17 NUMBER OF SEQUENCES: 14
18
19 CORRESPONDENCE ADDRESS:
20
21 ADDRESSEE: Genetics Institute, Inc.
22
23 STREET: 87 CambridgePark Drive
24
25 CITY: Cambridge
26
27 STATE: Massachusetts
28
29 COUNTRY: USA
30
31 ZIP: 02140
32
33 COMPUTER READABLE FORM:
34
35 MEDIUM TYPE: Floppy disk
36
37 COMPUTER: IBM PC compatible
38
39 OPERATING SYSTEM: PC-DOS/MS-DOS
40
41 SOFTWARE: PatentIn Release #1.0, Version #1.25
42
43 CURRENT APPLICATION DATA:
44
45 APPLICATION NUMBER: PCT/US95/12724
46
47 FILING DATE:

```

```

/
/
/ ATTORNEY/AGENT INFORMATION:
/
/ NAME: BROWN, SCOTT A,
/
/ REGISTRATION NUMBER: 32,724
/
/ REFERENCE/DOCKET NUMBER: G15232B
/
/ TELECOMMUNICATION INFORMATION:
/
/ TELEPHONE: (617) 498-8224
/
/ TELEFAX: (617) 876-5851
/
/ INFORMATION FOR SEQ ID NO: 1:
/
/ SEQUENCE CHARACTERISTICS:
/
/ LENGTH: 2158 base pairs
/
/ TYPE: nucleic acid
/
/ STRANDEDNESS: double
/
/ TOPOLOGY: linear
/
/ MOLECULE TYPE: cDNA
/
/ HYPOTHEICAL: NO
/
/ FEATURE:
/
/ NAME/KEY: CDS
/
/ LOCATION: 2..1231
/
/
/ PCT-US95-12724-1

```

```
Query Match      8.1%; Score 29; DB 6; Length 2158;
Best Local Similarity 53.0%; Pred. No. 1.8;
Matches 62: Conservative 0; Mismatches 55; Indels 0; Gaps 0;
```

195	tgtaagcaagatgagcaagaagcccaacggaagcggtcacattggctgtttaagccac	254
QY		
530	TGATTTCGTGGCTGTATACAAAGCCCAATGCGCACTTTCCCATTAGCGCGCTCACCTTCT	471
Db		
255	agctgttgcgctccagtcacacttctctaaaaacagggaagcactggagccagcgca	311
QY		
470	TGCGGATGTCATTCTTATTTACCTTCATCAGAGCATGTAGGAGATGAGGTTCTGCA	414
Db		

RESULT 11

US-08-698-551-11/c
; Sequence 11, Application US/08698551
; Patent No. 5712381
; GENERAL INFORMATION:
; APPLICANT: Lin, Lih-Ling
; APPLICANT: Chen, Jennifer H.
; APPLICANT: Schievella, Andrea
; APPLICANT: Graham, James
; TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
; TITLE OF INVENTION: PROTEINS AND INHIBITORS OF LIGAND BINDING
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.

; CURRENT APPLICATION DATA:

```
; LENGTH: 2735 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..1822
; US-08-494-440B-11

Query Match      8.1%; Score 29; DB 3; Length 2735;
Best Local Similarity 53.0%; Pred. No. 2;
Matches 62; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 195 tgttaagcaagatgagcagaagcccaacggaagcgtcacattggctgtttaagccaac 254
      || || || || || || || || || || || || || || || || || || || ||
DB 1121 TGATTTCGCTGTACACAAGCCCAATGTGCGACTTTCCCATTTAGGCGCCTCACCTTCT 1062

QY 255 agctgtggctccacgtaccacttctctctctctctctctctctctctctctctctct 311
      || || || || || || || || || || || || || || || || || || || ||
DB 1061 TCGGATGTCATCTTATTACCTTCATCAGCAGCATGTAGGATGAGGTTGTGCA 1005

RESULT 14
US-08-533-901B-11/c
; Sequence 11, Application US/08533901B
; Patent No. 5852173
; GENERAL INFORMATION:
; APPLICANT: Lin, Lih-Ling
; APPLICANT: Chen, Jennifer H.
; APPLICANT: Schievella, Andrea
; APPLICANT: Graham, James
; TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
; TITLE OF INVENTION: PROTEINS AND INHIBITORS OF LIGAND BINDING
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/533,901B
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; REFERENCE/DOCKET NUMBER: G15232
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2735 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..1822
; US-08-533-901B-11
```

```
Query Match      8.1%; Score 29; DB 3; Length 2735;
Best Local Similarity 53.0%; Pred. No. 2;
Matches 62; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 195 tgttaagcaagatgagcagaagcccaacggaagcgtcacattggctgtttaagccaac 254
      || || || || || || || || || || || || || || || || || || || ||
DB 1121 TGATTTCGCTGTACACAAGCCCAATGTGCGACTTTCCCATTTAGGCGCCTCACCTTCT 1062

QY 255 agctgtggctccacgtaccacttctctctctctctctctctctctctctctctctct 311
      || || || || || || || || || || || || || || || || || || || ||
DB 1061 TCGGATGTCATCTTATTACCTTCATCAGCAGCATGTAGGATGAGGTTGTGCA 1005

RESULT 15
US-08-839-032A-11/c
; Sequence 11, Application US/08839032A
; Patent No. 5891675
; GENERAL INFORMATION:
; APPLICANT: Lin, Lih-Ling
; APPLICANT: Chen, Jennifer H.
; APPLICANT: Schievella, Andrea
; APPLICANT: Graham, James
; TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND PROTEINS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/839,032A
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; REFERENCE/DOCKET NUMBER: G15232DDIV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2735 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..1822
; US-08-839-032A-11

Query Match      8.1%; Score 29; DB 3; Length 2735;
Best Local Similarity 53.0%; Pred. No. 2;
Matches 62; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 195 tgttaagcaagatgagcagaagcccaacggaagcgtcacattggctgtttaagccaac 254
      || || || || || || || || || || || || || || || || || || || ||
DB 1121 TGATTTCGCTGTACACAAGCCCAATGTGCGACTTTCCCATTTAGGCGCCTCACCTTCT 1062

QY 255 agctgtggctccacgtaccacttctctctctctctctctctctctctctctctctct 311
      || || || || || || || || || || || || || || || || || || || ||
DB 1061 TCGGATGTCATCTTATTACCTTCATCAGCAGCATGTAGGATGAGGTTGTGCA 1005
```

Search completed: December 18, 2000, 14:41:09
Job time: 16836 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 18, 2000, 10:00:33 ; Search time 38.84 Seconds
(without alignments)
3481.940 Million cell updates/sec

Title: US-09-301-906-4
Perfect score: 360
Sequence: 1 gtcagcggtcagtcagcgc.....aggcgccttgacgtttggc 360

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 480022 seqs, 187831343 residues
Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_36:*
1: /SID56/gcgdata/geneseq/geneseq/NA1980.DAT:*
2: /SID56/gcgdata/geneseq/geneseq/NA1981.DAT:*
3: /SID56/gcgdata/geneseq/geneseq/NA1982.DAT:*
4: /SID56/gcgdata/geneseq/geneseq/NA1983.DAT:*
5: /SID56/gcgdata/geneseq/geneseq/NA1984.DAT:*
6: /SID56/gcgdata/geneseq/geneseq/NA1985.DAT:*
7: /SID56/gcgdata/geneseq/geneseq/NA1986.DAT:*
8: /SID56/gcgdata/geneseq/geneseq/NA1987.DAT:*
9: /SID56/gcgdata/geneseq/geneseq/NA1988.DAT:*
10: /SID56/gcgdata/geneseq/geneseq/NA1989.DAT:*
11: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT:*
12: /SID56/gcgdata/geneseq/geneseq/NA1991.DAT:*
13: /SID56/gcgdata/geneseq/geneseq/NA1992.DAT:*
14: /SID56/gcgdata/geneseq/geneseq/NA1993.DAT:*
15: /SID56/gcgdata/geneseq/geneseq/NA1994.DAT:*
16: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT:*
17: /SID56/gcgdata/geneseq/geneseq/NA1996.DAT:*
18: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT:*
19: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT:*
20: /SID56/gcgdata/geneseq/geneseq/NA1999.DAT:*
21: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	360	100.0	360	21	GLRAV-3 polyprotei
2	360	100.0	6714	21	GLRAV-3 ORF 1a, en
3	360	100.0	17919	21	Grapevine leafroll
C 4	33	9.2	1178	14	AIGF gene #2. Hom
C 5	32.2	8.9	6714	21	GLRAV-3 ORF 1a, en
C 6	32.2	8.9	17919	21	Grapevine leafroll
C 7	31	8.6	997	14	AIGF gene #1. Hom
C 8	30.8	8.6	1149	21	Neisseria meningit
C 9	30.4	8.4	567	21	HIV codon altered
C 10	30.2	8.4	1105	20	Human prostate tum
C 11	29.8	8.3	360	21	GLRAV-3 polyprotei
C 12	29.4	8.2	807	16	T-cell receptor al

13	29.4	8.2	807	19	V01408
C 14	29.4	8.2	3385	21	295275
C 15	29.4	8.2	11808	19	V33433
C 16	29.4	8.2	11994	17	T43075
C 17	29.2	8.1	696	21	253906
C 18	29.2	8.1	696	21	253907
C 19	29.2	8.1	696	21	253908
C 20	29.2	8.1	849	21	253909
C 21	29.2	8.1	849	21	253910
C 22	29.2	8.1	849	21	253911
C 23	29.2	8.1	1149	21	253466
C 24	29.2	8.1	1149	21	253467
C 25	29.2	8.1	5059	20	X84332
C 26	29.2	8.1	5338	20	X13242
C 27	29	8.1	1560	16	Q83201
C 28	29	8.1	2158	17	T15228
C 29	29	8.1	2158	18	T94631
C 30	29	8.1	2735	17	T15233
C 31	29	8.1	2735	18	T94636
C 32	29	8.1	6002	18	T94638
C 33	29	8.1	7360	17	T43072
C 34	29	8.1	7360	19	V33410
C 35	28.6	7.9	910715	20	X20248
C 36	28.4	7.9	5360	20	X13155
C 37	28.2	7.8	300	21	A00658
C 38	28.2	7.8	864	7	M60488
C 39	28.2	7.8	864	11	Q03317
C 40	28.2	7.8	1094	19	T98728
C 41	28.2	7.8	6567	11	Q03324
C 42	28	7.8	435	19	V29023
C 43	28	7.8	463	18	X83249
C 44	28	7.8	463	19	V68847
C 45	28	7.8	1689	14	Q52498

ALIGNMENTS

RESULT 1	
Z49203	ID Z49203 standard; DNA; 360 BP.
XX	AC Z49203;
XX	AC Z49203;
DT	07-MAR-2000 (first entry)
DE	GLRAV-3 polyprotein proteinase domain DNA.
XX	XX
KW	Grapevine; leafroll; grapevine leafroll-associated virus 3; GLRAV-3;
KW	viral disease; yield loss; sugar content; inhibition; infection;
KW	replication; polyprotein; domain; proteinase; methyltransferase;
KW	helicase; RNA-dependent; RNA polymerase; untranslated region;
KW	transgenic plant; component; resistant; truncation; deletion; antisense;
XX	expression; detection; antibody; ds.
OS	Grapevine leafroll-associated virus 3.
XX	XX
FH	Key Location/Qualifiers
FT	mat_peptide 1..360
FT	/tag- a
FT	/product= "GLRAV-3 polyprotein proteinase domain"
XX	XX
PN	WO9955880-A1.
XX	XX
PD	04-NOV-1999.
XX	XX
PF	29-APR-1999; 99WO-US09307.
XX	XX
PR	29-APR-1998; 98US-0083404.
XX	XX
PA	(CORR) CORNELL RES FOUND INC.
PI	Gonsalves D, Ling K;

```

XX WPI; 2000-062035/05.
DR P-PSDB; Y58143.
XX
XX Newly isolated grapevine leafroll virus protein or polypeptide useful
XX for producing transgenic plants conferring viral disease resistance -
XX
XX Claim 12; Fig 5; 84pp; English.
XX
XX This sequence represents DNA encoding the proteinase domain of the
XX polyprotein (Y58148) from the grapevine leafroll-associated virus 3
XX (GLRaV-3). Leafroll is a serious viral disease, occurring wherever
XX grapes are grown. Although not lethal, it causes yield losses and
XX reduction in sugar content. The virus encodes several proteins,
XX which may serve as targets for the inhibition of viral infection or
XX replication. These proteins include the 242-248 kD polyprotein (Y58148),
XX encoded by open reading frame (ORF) 1a, and which comprises a proteinase
XX domain (Y58143), a methyltransferase domain (Y58144) and a helicase
XX domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF
XX 1b; and a protein of unspecified function (Y58147), encoded by ORF 11.
XX Nucleotides encoding these proteins, or fragments thereof, and the 5',
XX - and 3' untranslated regions (UTRs) of the genome (249201-249202) are
XX useful for the generation of transgenic plants and plant components.
XX Such transgenic plants may be resistant to viral disease, for example,
XX this property being conferred on the plants via the use of nucleotides
XX encoding truncated or internally deleted proteins, or via the use of
XX antisense nucleotides to inhibit viral gene expression. The nucleotides
XX may additionally be used in the detection of viral nucleic acids in a
XX tissue sample. The proteins can be used to generate antibodies which
XX may be used to detect GLRaV-3 in plant samples. The isolation of
XX GLRaV-3 facilitates the production of agents that reduce the risk of
XX infection or damage by the virus in vineyards.
XX
XX Sequence 360 BP; 90 A; 97 C; 100 G; 73 T; 0 other;

Query Match      100.0%; Score: 360; DB 21; Length 360;
Best Local Similarity 100.0%; Pred. No. 1.6e-112;
Matches 360; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtcagcggtcagtcagcggtgaggggagtgatggaagaggtcttgatgagggcaagg 60
Db 1 gtcagcggtcagtcagcggtgaggggagtgatggaagaggtcttgatgagggcaagg 60
QY 61 acctcaacttcgcaacttcgacgtgtctgatttcgacgtctgatttcgacgtgttct 120
Db 61 acctcaacttcgcaacttcgacgtgtctgatttcgacgtctgatttcgacgtgttct 120
QY 121 aatgcattacttgcgtacactacacgcgggtagtcgcgtatgcccgctcaagcgcgag 180
Db 121 aatgcattacttgcgtacactacacgcgggtagtcgcgtatgcccgctcaagcgcgag 180
QY 181 cagcctaaccggtgttaagcaagatgagcagaagcccaacgcaagcgtcacattgg 240
Db 181 cagcctaaccggtgttaagcaagatgagcagaagcccaacgcaagcgtcacattgg 240
QY 241 gctgttaagcacaacagctgttgctccacgtaccacttctctaaacacagaagcactg 300
Db 241 gctgttaagcacaacagctgttgctccacgtaccacttctctaaacacagaagcactg 300
QY 301 gaggcagcgaatcagtcaccacacagtcgtgtgagagagagcgccttgacgtttggc 360
Db 301 gaggcagcgaatcagtcaccacacagtcgtgtgagagagagcgccttgacgtttggc 360

RESULT 2
ID 249208
XX 249208 standard; DNA; 6714 BP.
AC 249208;
XX
DT 07-MAR-2000 (first entry)
XX

```

```

DE XX GLRaV-3 ORF 1a, encoding polyprotein.
KW XX Grapevine; leafroll; grapevine leafroll-associated virus 3; GLRaV-3;
KW viral disease; yield loss; sugar content; inhibition; infection;
KW replication; polyprotein; domain; proteinase; methyltransferase;
KW helicase; RNA-dependent; RNA polymerase; untranslated region;
KW transgenic plant; component; resistant; truncation; deletion; antisense;
KW expression; detection; antibody; ds.
XX
XX Grapevine leafroll-associated virus 3.
XX
XX Key Location/Qualifiers
XX CDS 1..6714
XX /tag= a
XX /note= "ORF 1a"
XX /product= "GLRaV-3 polyprotein"
XX mat_peptide 253..612
XX /tag= b
XX /product= "GLRaV-3 polyprotein proteinase domain
XX (Y58143)"
XX mat_peptide 1378..2193
XX /tag= c
XX /product= "GLRaV-3 polyprotein methyltransferase domain
XX (Y58144)"
XX mat_peptide 5764..6636
XX /tag= d
XX /product= "GLRaV-3 polyprotein helicase domain (Y58145)"
XX
XX W09955880-A1.
XX
XX 04-NOV-1999.
XX
XX 29-APR-1999; 99WO-US09307.
XX
XX 29-APR-1999; 98US-0083404.
XX
XX (CORR ) CORNELL RES FOUND INC.
XX
XX Gonsalves D, Ling K;
XX
XX WPI; 2000-062035/05.
XX P-PSDB; Y58148.
XX
XX Newly isolated grapevine leafroll virus protein or polypeptide useful
XX for producing transgenic plants conferring viral disease resistance -
XX
XX Claim 7; Fig 4; 84pp; English.
XX
XX This sequence represents open reading frame (ORF) 1a which encodes
XX the polyprotein of the grapevine leafroll-associated virus 3
XX (GLRaV-3). Leafroll is a serious viral disease, occurring wherever
XX grapes are grown. Although not lethal, it causes yield losses and
XX reduction in sugar content. The virus encodes several proteins,
XX which may serve as targets for the inhibition of viral infection or
XX replication. These proteins include the 242-248 kD polyprotein (Y58148),
XX encoded by open reading frame (ORF) 1a, and which comprises a proteinase
XX domain (Y58143), a methyltransferase domain (Y58144) and a helicase
XX domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF
XX 1b; and a protein of unspecified function (Y58147), encoded by ORF 11.
XX Nucleotides encoding these proteins, or fragments thereof, and the 5',
XX - and 3' untranslated regions (UTRs) of the genome (249201-249202) are
XX useful for the generation of transgenic plants and plant components.
XX Such transgenic plants may be resistant to viral disease, for example,
XX this property being conferred on the plants via the use of nucleotides
XX encoding truncated or internally deleted proteins, or via the use of
XX antisense nucleotides to inhibit viral gene expression. The nucleotides
XX may additionally be used in the detection of viral nucleic acids in a
XX tissue sample. The proteins can be used to generate antibodies which
XX may be used to detect GLRaV-3 in plant samples. The isolation of
XX GLRaV-3 facilitates the production of agents that reduce the risk of
XX infection or damage by the virus in vineyards.
XX
XX Sequence 6714 BP; 1675 A; 1429 C; 1877 G; 1733 T; 0 other;

```

Query Match		100.0%;	Score 360;	DB 21;	Length 6714;
Best Local Similarity		100.0%;	Pred. No. 6.1e-112;		
Matches 360;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	gtcagcgctcagtcagcgcgtgagagggggtgtaagaaggtcttgcagggcaag	60		
Db	253	gtcagcgctcagtcagcgcgtgagagggggtgtaagaaggtcttgcagggcaag	312		
QY	61	acctcaactccgaactccacgtgtctgtatttcagcgtgtattcgaagctgttct	120		
Db	313	acctcaactccgaactccacgtgtctgtatttcagcgtgtattcgaagctgttct	372		
QY	121	aatgcattacttgcgtacactaccacgggtagtcgctgtatgccccgcgtcaagcgcgag	180		
Db	373	aatgcattacttgcgtacactaccacgggtagtcgctgtatgccccgcgtcaagcgcgag	432		
QY	181	cagcctaaacccgctgttaagcaagatgagcagaagcccaaacggaagcggtcacattgg	240		
Db	433	cagcctaaacccgctgttaagcaagatgagcagaagcccaaacggaagcggtcacattgg	492		
QY	241	gctgttaagccaacagctgttgccgtccacgtaccacttccctaaaaaacagggaagcactg	300		
Db	493	gctgttaagccaacagctgttgccgtccacgtaccacttccctaaaaaacagggaagcactg	552		
QY	301	gagccagcgaatcagtcgcccaacacagctgttgaggagaaagccgcttgacgtttggc	360		
Db	553	gagccagcgaatcagtcgcccaacacagctgttgaggagaaagccgcttgacgtttggc	612		
RESULT 3					
249200		standard; DNA; 17919 BP.			
XX	AC	249200;			
XX	DT	07-MAR-2000 (first entry)			
XX	DE	Grapevine leafroll-associated virus 3 (GLRaV-3) genome.			
KW	KW	Grapevine; leafroll; grapevine leafroll-associated virus 3; GLRaV-3;			
KW	KW	viral disease; yield loss; sugar content; inhibition; infection;			
KW	KW	replication; polyprotein; domain; proteinase; methyltransferase;			
KW	KW	helicase; RNA-dependent; RNA polymerase; untranslated region;			
KW	KW	transgenic plant; component; resistant; truncation; deletion; antisense;			
KW	KW	expression; detection; antibody; ds.			
XX	OS	Grapevine leafroll-associated virus 3.			
XX	XX	Key Location/Qualifiers			
FT	5'UTR	1..158			
FT	·	/*tag= a			
FT	·	159..6872			
FT	·	/*tag= b			
FT	·	/note= "ORF 1a"			
FT	·	/product= "GLRaV-3 polyprotein (Y58148)"			
FT	·	411..770			
FT	·	/*tag= c			
FT	·	/product= "GLRaV-3 polyprotein proteinase domain (Y58143)"			
FT	·	1536..2351			
FT	·	/*tag= d			
FT	·	/product= "GLRaV-3 polyprotein methyltransferase domain (Y58144)"			
FT	·	5922..6794			
FT	·	/*tag= e			
FT	·	/product= "GLRaV-3 polyprotein helicase domain (Y58145)"			
FT	·	6877..8478			
FT	·	/*tag= f			
FT	·	/note= "ORF 1b"			
FT	·	/product= "GLRaV-3 RNA-dependent RNA polymerase (Y58146)"			
FT	·	8708..8863			
FT	CDS				

FT	/*tag= g	
FT	/note= "ORF 2"	
FT	9930..10067	
FT	/*tag= h	
FT	/note= "ORF 3"	
FT	10086..11735	
FT	/*tag= i	
FT	/note= "ORF 4"	
FT	11728..13179	
FT	/*tag= j	
FT	/note= "ORF 5"	
FT	13269..14210	
FT	/*tag= k	
FT	/note= "ORF 6"	
FT	14273..15706	
FT	/*tag= l	
FT	/note= "ORF 7"	
FT	15717..16274	
FT	/*tag= m	
FT	/note= "ORF 8"	
FT	16271..16804	
FT	/*tag= n	
FT	/note= "ORF 9"	
FT	16811..17350	
FT	/*tag= o	
FT	/note= "ORF 10"	
FT	17353..17463	
FT	/*tag= p	
FT	/note= "ORF 11"	
FT	/product= "GLRaV-3 ORF 11-encoded protein (Y58147)"	
FT	17460..17642	
FT	/*tag= q	
FT	/note= "ORF 12"	
FT	17643..17919	
FT	/*tag= r	
XX	WO9955880-A1.	
XX	04-NOV-1999.	
XX	29-APR-1999;	99WO-US09307.
XX	29-APR-1998;	98US-0083404.
XX	(CORR) CORNELL RES FOUND INC.	
PI	Gonsalves D, Ling K;	
XX	WPI; 2000-062035/05.	
DR	P-PSDB; Y58143, Y58144, Y58145, Y58146, Y58147, Y58148.	
XX	Newly isolated grapevine leafroll virus protein or polypeptide useful for producing transgenic plants conferring viral disease resistance -	
PT	Example 1; Fig 2; 84pp; English.	
XX	This sequence represents the genome of grapevine leafroll-associated virus 3 (GLRaV-3). Leafroll is a serious viral disease, occurring wherever grapes are grown. Although not lethal, it causes yield losses and reduction in sugar content. The virus encodes several proteins, which may serve as targets for the inhibition of viral infection or replication. These proteins include a 242-248 kD polyprotein (Y58148), encoded by open reading frame (ORF) 1a, and which comprises a proteinase domain (Y58143), a methyltransferase domain (Y58144) and a helicase domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF 1b; and a protein of unspecified function (Y58147), encoded by ORF 11. Nucleotides encoding these proteins, or fragments thereof, and the 5' and 3' untranslated regions (UTRs) of the genome (249201-249202) are useful for the generation of transgenic plants and plant components. Such transgenic plants may be resistant to viral disease, for example, this property being conferred on the plants via the use of nucleotides encoding truncated or internally deleted proteins, or via the use of antisense nucleotides to inhibit viral gene expression. The nucleotides	


```

XX PF 29-APR-1999; 99WO-US09307.
XX KW replication; domain; proteinase; methyltransferase;
XX KW helicase; RNA-dependent; RNA polymerase; untranslated region;
XX KW transgenic plant; component; resistant; truncation; deletion; antisense;
XX KW expression; detection; antibody; ds.
XX OS Grapevine leafroll-associated virus 3.
XX XX
XX FH Key Location/Qualifiers
XX FT 5'UTR 1..158
XX FT FT /*tag= a
XX FT CDS 159..6872
XX FT FT /*tag= b
XX FT FT /note= "ORF 1a"
XX FT FT /product= "GLRaV-3 polyprotein (Y58148)"
XX FT mat_peptide 411..770
XX FT FT /*tag= c
XX FT FT /product= "GLRaV-3 polyprotein proteinase domain
XX FT (Y58143)"
XX FT mat_peptide 1536..2351
XX FT FT /*tag= d
XX FT FT /product= "GLRaV-3 polyprotein methyltransferase domain
XX FT (Y58144)"
XX FT mat_peptide 5922..6794
XX FT FT /*tag= e
XX FT FT /product= "GLRaV-3 polyprotein helicase domain (Y58145)"
XX FT CDS 6877..8478
XX FT FT /*tag= f
XX FT FT /note= "ORF 1b"
XX FT FT /product= "GLRaV-3 RNA-dependent RNA polymerase (Y58146)"
XX FT CDS 8708..8863
XX FT FT /*tag= g
XX FT FT /note= "ORF 2"
XX FT CDS 9930..10067
XX FT FT /*tag= h
XX FT FT /note= "ORF 3"
XX FT CDS 10086..11735
XX FT FT /*tag= i
XX FT FT /note= "ORF 4"
XX FT CDS 11728..13179
XX FT FT /*tag= j
XX FT FT /note= "ORF 5"
XX FT CDS 13269..14210
XX FT FT /*tag= k
XX FT CDS 14273..15706
XX FT FT /*tag= l
XX FT CDS 15717..16274
XX FT FT /*tag= m
XX FT CDS 16271..16804
XX FT FT /*tag= n
XX FT CDS 16811..17350
XX FT FT /*tag= o
XX FT CDS 17353..17463
XX FT FT /*tag= p
XX FT FT /note= "ORF 11"
XX FT FT /product= "GLRaV-3 ORF 11-encoded protein (Y58147)"
XX FT CDS 17460..17642
XX FT FT /*tag= q
XX FT FT /note= "ORF 12"
XX FT 3'UTR 17643..17919
XX FT FT /*tag= r
XX XX
XX PN WO955880-A1.
XX XX
XX PD 04-NOV-1999.
XX XX
XX PF 29-APR-1999; 99WO-US09307.
XX XX
XX PR 29-APR-1998; 98US-0083404.
XX PR

```

This sequence represents open reading frame (ORF) 1a which encodes the polyprotein of the grapevine leafroll-associated virus 3 (GLRaV-3). Leafroll is a serious viral disease, occurring wherever grapes are grown. Although not lethal, it causes yield losses and reduction in sugar content. The virus encodes several proteins, CC which may serve as targets for the inhibition of viral infection or CC replication. These proteins include the 242-248 kD polyprotein (Y58148), CC encoded by open reading frame (ORF) 1a, and which comprises a proteinase CC domain (Y58143), a methyltransferase domain (Y58144), and a helicase CC domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF CC 1b; and a protein of unspecified function (Y58147), encoded by ORF 11. CC Nucleotides encoding these proteins, or fragments thereof, and the 5' CC and 3' untranslated regions (UTRs) of the genome (249201-249202) are CC useful for the generation of transgenic plants and plant components. CC Such transgenic plants may be resistant to viral disease, for example, CC this property being conferred on the plants via the use of nucleotides CC encoding truncated or internally deleted proteins, or via the use of CC antisense nucleotides to inhibit viral gene expression. The nucleotides CC may additionally be used in the detection of viral nucleic acids in a CC tissue sample. The proteins can be used to generate antibodies which CC may be used to detect GLRaV-3 in plant samples. The isolation of CC GLRaV-3 facilitates the production of agents that reduce the risk of CC infection or damage by the virus in vineyards.

Query Match 8.9%; Score 32.2; DB 21; Length 6714;
 Best Local Similarity 51.8%; Pred. No. 0.88;
 Matches 73; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

```

Qy 220 aaagcggaagctcacattggcgtgtaagccacacagctgttgccgctaccactt 279
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 616 AAAGGCCAAACGTCAGGGCGGCTTCTCCCAACGACTGTTGGGACTGATGGCTG 557
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 280 cctaaaaaacagggaagcactggagccagcgcaatcagtcacacagctgttgaggag 339
    || || || || || || || || || || || || || || || || || || ||
Db 556 GCTCCAGTGCTCTCTGTTTTAGGAAGTGGTACGTGAGCGCCAAACAGCTGTGGCTTAA 497
    || || || || || || || || || || || || || || || || || || ||
Qy 340 aagggccgcttgacgtttggc 360
    || || || || || || || || || || || || || || || || || || ||
Db 496 CAGCCCAATGTGACGCTGCC 476
    || || || || || || || || || || || || || || || || || || ||

```

RESULT 6
 249200/c
 ID 249200 standard; DNA; 17919 BP.
 XX AC 249200;
 XX XX
 XX DT 07-MAR-2000 (first entry)
 XX DE
 XX DE Grapevine leafroll-associated virus 3 (GLRaV-3) genome.
 XX KW Grapevine; leafroll; grapevine leafroll-associated virus 3; GLRaV-3;
 KW viral disease; yield loss; sugar content; inhibition; infection;

```

KW replication; domain; proteinase; methyltransferase;
KW helicase; RNA-dependent; RNA polymerase; untranslated region;
KW transgenic plant; component; resistant; truncation; deletion; antisense;
KW expression; detection; antibody; ds.
OS Grapevine leafroll-associated virus 3.
XX XX
XX FH Key Location/Qualifiers
XX FT 5'UTR 1..158
XX FT FT /*tag= a
XX FT CDS 159..6872
XX FT FT /*tag= b
XX FT FT /note= "ORF 1a"
XX FT FT /product= "GLRaV-3 polyprotein (Y58148)"
XX FT mat_peptide 411..770
XX FT FT /*tag= c
XX FT FT /product= "GLRaV-3 polyprotein proteinase domain
XX FT (Y58143)"
XX FT mat_peptide 1536..2351
XX FT FT /*tag= d
XX FT FT /product= "GLRaV-3 polyprotein methyltransferase domain
XX FT (Y58144)"
XX FT mat_peptide 5922..6794
XX FT FT /*tag= e
XX FT FT /product= "GLRaV-3 polyprotein helicase domain (Y58145)"
XX FT CDS 6877..8478
XX FT FT /*tag= f
XX FT FT /note= "ORF 1b"
XX FT FT /product= "GLRaV-3 RNA-dependent RNA polymerase (Y58146)"
XX FT CDS 8708..8863
XX FT FT /*tag= g
XX FT FT /note= "ORF 2"
XX FT CDS 9930..10067
XX FT FT /*tag= h
XX FT FT /note= "ORF 3"
XX FT CDS 10086..11735
XX FT FT /*tag= i
XX FT FT /note= "ORF 4"
XX FT CDS 11728..13179
XX FT FT /*tag= j
XX FT FT /note= "ORF 5"
XX FT CDS 13269..14210
XX FT FT /*tag= k
XX FT CDS 14273..15706
XX FT FT /*tag= l
XX FT CDS 15717..16274
XX FT FT /*tag= m
XX FT CDS 16271..16804
XX FT FT /*tag= n
XX FT CDS 16811..17350
XX FT FT /*tag= o
XX FT CDS 17353..17463
XX FT FT /*tag= p
XX FT FT /note= "ORF 11"
XX FT FT /product= "GLRaV-3 ORF 11-encoded protein (Y58147)"
XX FT CDS 17460..17642
XX FT FT /*tag= q
XX FT FT /note= "ORF 12"
XX FT 3'UTR 17643..17919
XX FT FT /*tag= r
XX XX
XX PN WO955880-A1.
XX XX
XX PD 04-NOV-1999.
XX XX
XX PF 29-APR-1999; 99WO-US09307.
XX XX
XX PR 29-APR-1998; 98US-0083404.
XX PR

```

(CORR) CORNELL RES FOUND INC.
Gonsalves D, Ling K;
WPI: 2000-062035/05.
P-PSDB: Y58143, Y58144, Y58145, Y58146, Y58147, Y58148.
Newly isolated grapevine leafroll virus protein or polypeptide useful
for producing transgenic plants conferring viral disease resistance -
Example 1; Fig 2; 84pp; English.
This sequence represents the genome of grapevine leafroll-associated
virus 3 (GLRaV-3). Leafroll is a serious viral disease, occurring
wherever grapes are grown. Although not lethal, it causes yield losses
and reduction in sugar content. The virus encodes several proteins,
which may serve as targets for the inhibition of viral infection or
replication. These proteins include a 242-248 kD polyprotein (Y58148),
domain (Y58143), a methyltransferase domain (Y58144) and a helicase
domain (Y58145); a RNA-dependent RNA polymerase (Y58146) and a helicase
lb; and a protein of unspecified function (Y58147), encoded by ORF 11.
Nucleotides encoding these proteins, or fragments thereof, and the 5'
and 3' untranslated regions (UTRs) of the genome (Z49201-Z49202) are
useful for the generation of transgenic plants and plant components.
Such transgenic plants may be resistant to viral disease, for example,
this property being conferred on the plants via the use of nucleotides
encoding truncated or internally deleted proteins, or via the use of
antisense nucleotides to inhibit viral gene expression. The nucleotides
may additionally be used in the detection of viral nucleic acids in a
tissue sample. The proteins can be used to generate antibodies which
can be used to detect GLRaV-3 in plant samples. The isolation of
GLRaV-3 facilitates the production of agents that reduce the risk of
infection or damage by the virus in vineyards.
Sequence 17919 BP: 4729 A; 3527 C; 4740 G; 4923 T; 0 other;
Query Match 8.9%; Score 32.2; DB 21; Length 17919;
Best Local Similarity 51.8%; Pred. No. 1.4;
Matches 73; Conservative 0; Mismatches 68; Indels 0; Gaps
QY 220 aaacgcgaacggtccacattggcgtttaagcacaacagctgttggcgctcacgtaccatt 279
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 774 AAAGGCCAAACGTCAGCGCGCTTCTCTCCCAACGACTGTTCTGGGACTGATTCGGCTG 715
QY 280 cctaaacaaacgaagacactgagcgcgcgaatcagtcaccaacacagtcgcttgagag 339
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 714 GCTCCAGTGCCTCCTGTTTATAGGAAGTGGTACGTGGACGCCAACAGCTGTGGCTTAA 655
QY 340 aadgcgcgcctgaactgttggc 360
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 654 CAGCCCAATGTGACGCTTGCC 634
RESULT 7
Q48024/c
ID Q48024 standard; DNA: 997 BP.
XX Q48024:
XX
XX 21-APR-1994 (first entry)
XX
XX AIGF gene #1.
XX
XX Cellular growth factor; androgen-induced growth factor; AIGF;
KW cell growth; reagent; detection; anticancer; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FT CDS 174..821

PR	02-SEP-1998;	98US-0099062.	
PR	09-OCT-1998;	98US-0103749.	
PR	09-OCT-1998;	98US-0103794.	
PR	09-OCT-1998;	98US-0103796.	
PR	09-FEB-1999;	98US-0121528.	
XX			
XX	(CHIR) CHIRON CORP.		
PA	(GENO-) INST GENOMIC RES.		
PA			
XX			
XX	Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;		
PI	Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;		
PI	Tettelin H, Venter JC;		
PI			
XX			
XX	WPI: 2000-062150/05.		
DR	P-PSDB; Y74706.		
XX			
XX	Novel Neisserial polypeptides predicted to be useful antigens for		
PT	vaccines and diagnostics		
PT			
XX	Claim 7: Page 542; 1453pp; English.		
PS			
XX			
XX	Z53015 to Z54536, Z54577 to Z54615, and Y74253 to Y75941 represent		
CC	novel <i>Neisseria meningitis</i> and <i>N. gonorrhoeae</i> polynucleotides and		
CC	polypeptides. Z54537 to Z54576 and Z54616 to Z55473 represent PCR		
CC	primers used in the exemplification of the present invention. The		
CC	polypeptides, the polynucleotides, antibodies and compositions of		
CC	the invention can be used as vaccines, as diagnostic reagents, and as		
CC	immunogenic compositions. The polypeptides can be used in the		
CC	manufacture of medicaments for treating or preventing infection due to		
CC	<i>Neisserial bacteria</i> (e.g. meningitis and septicemia), to detect the		
CC	presence of <i>Neisseria bacteria</i> , or to raise antibodies. They may also		
CC	be used to screen for agonists or antagonists, which may themselves		
CC	have use as antibacterial agents. The polynucleotides of the invention		
CC	may also be used in gene therapy protocols.		
XX			
XX	Sequence 1149 BP; 288 A; 297 C; 286 G; 278 T; 0 other;		
SQ			

	Query Match	8.5%	Score 30.8;	DB 21;	Length 1149;
	Best Local Similarity	70.7%;	Pred. No. 1.1;		
	Matches 41;	Conservative	0;	Mismatches 17;	Indels 0;
	Gaps	0;			
QY . 282	tataaacacggaagcactgagcgacgcgaatcagtcgccacaacagctcggttgaggag	339			
Db	802 TGAATAACAGTCCGTTTGAGCAAGCGCAATGCTTTTGGCAACAGCTCTTTTCAGGCG	745			

RESULT	9	
A29550		
ID	A29550 standard; DNA; 567 BP.	
XX		
XX	A29550;	
AC		
XX	09-AUG-2000 (first entry)	
DT		
XX		
XX	HIV codon altered env nucleotide sequence #2.	
DE		
XX		
KW	Erythropoietin; EPO; G-CSF; granulocyte colony stimulating factor;	
KW	wobble; codon altered gene; shuffling; modification; vaccine;	
KW	insulin; peptide hormone; growth factor; cytokine; interferon;	
KW	interleukin; leukaemia inhibitory factor; oncostatin M;	
KW	transcription activator; expression activator; infectious organism; ds.	
XX		
XX	Human immunodeficiency virus type 1.	
OS	Synthetic.	
OS		
XX	WO200018906-A2.	
PN		
XX	06-APR-2000.	
PN		
PD		
XX	28-SEP-1999; 99WO-US22588.	
PF		
XX		
XX	29-SEP-1998; 98US-0102362.	
PR		

[illegible]

```
XX 28-APR-1998; 98DE-1020190.
XX (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
XX Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
XX WPI; 1999-621386/54.
XX P-PSDB; Y73950, Y73951, Y73952.
XX
XX New human nucleic acid sequences from pancreatic tumors, and related
XX proteins
XX
XX Claim 2; Page 219-220; 502pp; German.
XX
XX This invention describes novel polypeptides and their encoding nucleic
XX acids derived from human pancreatic tumor tissue which have cytostatic
XX activity. The sequences are also useful in producing pharmaceutical
XX compositions for treatment of pancreatic tumors. 252858-253014 represent
XX expressed sequence tag (EST) fragments derived from a human pancreatic
XX tumor cDNA library and which encode the proteins represented in
XX . Y73814-Y74252.
XX
XX Sequence 1105 BP; 275 A; 299 C; 328 G; 203 T; 0 other;

Query Match      8.4%; Score 30.2; DB 20; Length 1105;
Best Local Similarity 53.9%; Pred. No. 1.8;
Matches 62; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

Qy 106 ttccgaagctgttttaagtcattactgtcgtacactaccacgggtagtgcgtatgcc 165
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1100 TTCAAACAGGTAACCTAATAAGCTGTGCGCCCTGCGCCGCCACCGTGAATACGCC 1041

Qy 166 ccgcgtacagcgagcagcctaaacccgctgttaagcaagatgacagaagccca 220
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1040 CCAGTGGGTGAACACAGGCCCACTGTGGGTATTGATGAACATGGGCACCAAGCCCA 986

RESULT 11
ID 249203 standard; DNA; 360 BP.
AC 249203;
DT 07-MAR-2000 (first entry)
XX GLRav-3 polyprotein proteinase domain DNA.
XX
XX Grapevine; leafroll; grapevine leafroll-associated virus 3; GLRav-3;
XX viral disease; yield loss; sugar content; inhibition; infection;
XX replication; polyprotein; domain; proteinase; methyltransferase;
XX helicase; RNA-dependent; RNA polymerase; untranslated region;
XX transgenic plant; component; resistant; truncation; deletion; antisense;
XX expression; detection; antibody; ds.
XX
XX Grapevine leafroll-associated virus 3.
XX
XX Key Location/Qualifiers
XX mat_peptide 1..360
XX /*tag= a
XX /product= "GLRav-3 polyprotein proteinase domain"
XX
XX WO9955880-A1.
XX
XX 04-NOV-1999.
XX
XX 29-APR-1999; 99WO-US09307.
XX
XX 29-APR-1998; 98US-0083404.
XX (CORR ) CORNELL RES FOUND INC.
XX
```

```
PI Gonsalves D, Ling K;
XX
XX WPI; 2000-062035/05.
XX P-PSDB; Y58143.
XX
XX Newly isolated grapevine leafroll virus protein or polypeptide useful
XX for producing transgenic plants conferring viral disease resistance -
XX
XX Claim 12; Fig 5; 84pp; English.
XX
XX This sequence represents DNA encoding the proteinase domain of the
XX polyprotein (Y58148) from the grapevine leafroll-associated virus 3
XX (GLRav-3). Leafroll is a serious viral disease, occurring wherever
XX grapes are grown. Although not lethal, it causes yield losses and
XX reduction in sugar content. The virus encodes several proteins,
XX which may serve as targets for the inhibition of viral infection or
XX replication. These proteins include the 242-248 kD polyprotein (Y58148),
XX encoded by open reading frame (ORF) 1a, and which comprises a proteinase
XX domain (Y58143), a methyltransferase domain (Y58144) and a helicase
XX domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF
XX 1b; and a protein of unspecified function (Y58147), encoded by ORF 11.
XX Nucleotides encoding these proteins, or fragments thereof, and the 5',
XX and 3' untranslated regions (UTRs) of the genome (249201-249202) are
XX useful for the generation of transgenic plants and plant components.
XX Such transgenic plants may be resistant to viral disease, for example,
XX this property being conferred on the plants via the use of nucleotides
XX encoding truncated or internally deleted proteins, or via the use of
XX antisense nucleotides to inhibit viral gene expression. The nucleotides
XX may additionally be used in the detection of viral nucleic acids in a
XX tissue sample. The proteins can be used to generate antibodies which
XX may be used to detect GLRav-3 in plant samples. The isolation of
XX GLRav-3 facilitates the production of agents that reduce the risk of
XX infection or damage by the virus in vineyards.
XX
XX Sequence 360 BP; 90 A; 97 C; 100 G; 73 T; 0 other;

Query Match      8.3%; Score 29.8; DB 21; Length 360;
Best Local Similarity 51.1%; Pred. No. 1.5;
Matches 70; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

Qy 224 gccaagcgtcacattggcgtgttaagccaacagctgtgtgctccacgtaccacttcta 283
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 360 GCCAAACGTCAGGCGCGCTTCTCTCCCAACGACTGTGTGGGACTGATTGGCGTGGCTC 301

Qy 284 aaaaacaggagcactggagccagcgcaatcagtcaccaacagctgttgagagaagg 343
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 300 CAGTGTCTTCTGTTTATTAGGAAGTGTGCTGACGACGCCAACAGCTGTGGCTTACAGC 241

Qy 344 ccgccttgacgtttggc 360
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 240 CCAATGTGACGCTTGCC 224

RESULT 12
Q91362
ID Q91362 standard; cDNA; 807 BP.
XX
XX AC Q91362;
XX
XX DT 24-NOV-1995 (first entry)
XX
XX DE T-cell receptor alpha chain (TCR alpha) from hybridoma 3B3.
XX
XX KW T-cell receptor alpha chain; TCR alpha; hyperimmune;
XX immunodeficiency; Cd4+; helper T cell; hybridoma 3B3; ss.
XX
XX OS Synthetic.
XX
XX Key Location/Qualifiers
XX CDS 1..804
XX /*tag= a
XX
```



```
FT      /product= "IKR-1"
FT      /note= "I kappa B kinase-related kinase 1"
PN      WO200008179-A1.
XX      17-FEB-2000.
PD      04-AUG-1999; 99WO-US17578.
PF      04-AUG-1998; 98US-0095269.
PR      11-SEP-1998; 98US-0099973.
PR      05-FEB-1999; 99US-0118783.
XX      (IMV ) IMMUNEX CORP.
PA      Bird TA, Virca GD;
XX      WPI; 2000-195583/17.
DR      P-PSDB; Y80279.
XX      Novel kappa B-kinase related kinases IKR-1 and IKR-2 used as molecular
PT      weight markers and in peptide fragmentation studies
PS      Claim 1; Fig 1; 85pp; English.
XX      The present sequence encodes murine I kappa B-kinase related kinase 1
CC      (IKR-1). IKR proteins have immunomodulatory, antiinflammatory,
CC      antimicrobial and cytostatic activities. IKR polynucleotides can be
CC      used to express the proteins, and as probes to identify nucleic acids
CC      encoding proteins having kinase activity. IKR-1 and IKR-2 proteins and
CC      fragmented polypeptides are used for purifying proteins, e.g. to purify
CC      binding partner proteins; to measure protein activity, e.g. as quality
CC      assurance agents to monitor shelf life and stability of binding partner
CC      proteins. They may also be used as research agents, e.g. in assays to
CC      determine protein kinase activity, to identify novel molecules involved
CC      in signal transduction pathways, and to identify therapeutic compounds,
CC      to identify substances which interfere with the rate of substrate
CC      phosphorylation (such compounds would be useful for the treatment of
CC      autoimmune, inflammatory, infectious or neoplastic diseases), as
CC      molecular weight and isoelectric focusing markers, as controls for
CC      peptide fragmentation, identification of unknown proteins, e.g. by
CC      comparison with proteins in databases and for preparation of antibodies.
CC      The antibodies can be used in assays to detect the presence of the
CC      protein, and to purify the protein by immunoaffinity chromatography. The
CC      antibodies can also be used to block binding of the IKR polypeptides to
CC      their binding partners.
XX      Sequence 3385 BP; 821 A; 922 C; 925 G; 717 T; 0 other;
SQ      Query Match      8.2%; Score 29.4; DB 21; Length 3385;
        Best Local Similarity 58.6%; Pred. No. 5.7;
        Matches 51; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY      225 gcaagcgtcacattgggtgtaagccacagctgttggtccacgtaccacttcctaa 284
        ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      3083 GAAAAGGTCAAGGAGATCCTCAAGACACAGCAGTAGCTCAACCCACCACCTTCTCA 3024
        ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      285 aaacaggaagcactggagccagcga 311
        ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      3023 AATGCAGGTGATCTAGACCAAGCACA 2997
        ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
V33433/c
ID      V33433 standard; DNA; 11808 BP.
XX
AC      V33433;
XX
DT      18-NOV-1998 (first entry)
XX
DE      Triticum aestivum acetyl-CoA carboxylase (ACC) genomic DNA sequence.
XX
```

```
KW      Triticum aestivum acetyl-CoA carboxylase; ACC; plant oil;
KW      herbicide resistance; ss.
XX
OS      Triticum aestivum.
XX
PN      US5801233-A.
XX
PD      01-SEP-1998.
XX
PF      05-MAR-1996; 96US-0611107.
XX
PR      05-MAR-1996; 96US-0611107.
PR      02-OCT-1992; 92US-0956700.
PR      14-APR-1995; 95US-0422560.
XX
PA      (ARCH-) ARCH DEV CORP.
XX
PI      Gornicki P, Haselkorn R;
XX      WPI; 1998-494841/42.
DR      P-PSDB; W70409.
XX
PT      Isolated DNA encoding plant and cyanobacterial acetyl-CoA
PT      carboxylase polypeptides - useful for producing recombinant
PT      polypeptides and increasing the herbicide resistance of plants
PS      Claim 3; Fig 4A-4I; 96pp; English.
XX
CC      The present claimed sequence represents the Triticum aestivum
CC      acetyl-CoA carboxylase (ACC) genomic DNA sequence. The invention
CC      claims for novel ACC proteins and the DNA encoding these proteins
CC      from plant (e.g. wheat and canola) and cyanobacterial (e.g. Anabaena
CC      and Synechococcus) species. The ACCs of the invention are claimed
CC      to be useful for regulating the oil content of plant tissues, for
CC      conferring and modulating herbicide resistance in plants, and for
CC      altering the activity of ACC in plant cells in vivo.
XX
SQ      Sequence 11808 BP; 3107 A; 2467 C; 2695 G; 3539 T; 0 other;

Query Match      8.2%; Score 29.4; DB 19; Length 11808;
Best Local Similarity 63.4%; Pred. No. 10;
Matches 45; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY      57 aaggaccctcaacttcgcgaacttcgcgaacttcgcgaacttcgcgaacttcgcga 116
        ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      9034 ATGCTCATTAGATTCTGCAACATGTGAATACACAGATTCCTCAATGCTGT 8975
        ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      117 ttctaattgcac 127
        ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      8974 TTCAATGCTT 8964

Search completed: December 18, 2000, 14:40:36
Job time: 16803 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 18, 2000, 10:00:33 ; Search time 856.91 Seconds
(without alignments)
2597.484 Million cell updates/sec

Title: US-09-301-906-4
Perfect score: 360

Sequence: 1 gtcagcggtcagtcagcgc.....aggccgcttgacgtttggc 360

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
- 1: gb_est1.*
 - 2: gb_est2.*
 - 3: gb_est3.*
 - 4: gb_est4.*
 - 5: gb_est5.*
 - 6: gb_est6.*
 - 7: gb_est7.*
 - 8: gb_est8.*
 - 9: gb_est9.*
 - 10: gb_est10.*
 - 11: gb_est11.*
 - 12: gb_est12.*
 - 13: gb_est13.*
 - 14: gb_est14.*
 - 15: gb_est15.*
 - 16: gb_est16.*
 - 17: gb_est17.*
 - 18: gb_est18.*
 - 19: gb_est19.*
 - 20: gb_est20.*
 - 21: gb_est21.*
 - 22: gb_est22.*
 - 23: gb_est23.*
 - 24: gb_est24.*
 - 25: gb_est25.*
 - 26: gb_est26.*
 - 27: gb_est27.*
 - 28: gb_est28.*
 - 29: gb_est29.*
 - 30: gb_est30.*
 - 31: gb_est31.*
 - 32: gb_est32.*
 - 33: gb_est33.*
 - 34: gb_est34.*
 - 35: gb_est35.*
 - 36: gb_est36.*
 - 37: gb_est37.*
 - 38: gb_est38.*
 - 39: gb_est39.*
 - 40: gb_est40.*
 - 41: em_estba.*
 - 42: em_estfun.*
 - 43: em_esthum1.*

- 44: em_esthum2.*
- 45: em_esthum3.*
- 46: em_esthum4.*
- 47: em_esthum5.*
- 48: em_esthum6.*
- 49: em_esthum7.*
- 50: em_esthum8.*
- 51: em_esthum9.*
- 52: em_esthum10.*
- 53: em_esthum11.*
- 54: em_esthum12.*
- 55: em_esthum13.*
- 56: em_esthum14.*
- 57: em_esthum15.*
- 58: em_esthum16.*
- 59: em_esthum17.*
- 60: em_esthum18.*
- 61: em_esthum19.*
- 62: em_esthum20.*
- 63: em_estin1.*
- 64: em_estin2.*
- 65: em_estin3.*
- 66: em_estin4.*
- 67: em_estom.*
- 68: em_estov1.*
- 69: em_estov2.*
- 70: em_estpl1.*
- 71: em_estpl2.*
- 72: em_estpl3.*
- 73: em_estpl4.*
- 74: em_estpl5.*
- 75: em_estrol.*
- 76: em_estro2.*
- 77: em_estro3.*
- 78: em_estro4.*
- 79: em_estro5.*
- 80: em_estro6.*
- 81: em_estro7.*
- 82: em_estro8.*
- 83: em_estro9.*
- 84: em_estrol1.*
- 85: em_estrol2.*
- 86: em_estrol3.*
- 87: gb_gss1.*
- 88: gb_gss2.*
- 89: gb_gss3.*
- 90: gb_gss4.*
- 91: gb_gss5.*
- 92: gb_gss6.*
- 93: gb_gss7.*
- 94: gb_gss8.*
- 95: gb_gss9.*
- 96: gb_gss10.*
- 97: gb_gss11.*
- 98: gb_gss12.*
- 99: gb_gss13.*
- 100: gb_gss14.*
- 101: gb_gss15.*
- 102: gb_gss16.*
- 103: gb_gss17.*
- 104: gb_gss18.*
- 105: gb_gss19.*
- 106: gb_gss20.*
- 107: gb_gss21.*
- 108: gb_gss22.*
- 109: gb_gss23.*
- 110: gb_gss24.*
- 111: gb_gss25.*
- 112: gb_gss26.*
- 113: gb_gss27.*
- 114: gb_gss28.*
- 115: gb_gss29.*
- 116: gb_gss30.*

117: gb_gss18:*
 118: gb_gss19:*
 119: em_gss13:*
 120: gb_gss20:*
 121: gb_gss21:*
 122: gb_gss22:*
 123: gb_gss23:*
 124: gb_gss24:*
 125: em_gss14:*
 126: em_gss15:*
 127: em_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	36.8	10.2	276	27	BB155217	BB155217
C 2	36.4	10.1	370	40	W99276	W99276 T3079 WVA74
C 3	36.4	10.1	409	40	W69025	W69025 T2568 WVA74
C 4	36.4	10.1	661	100	AQ651482	AQ651482 Sheared D
C 5	35.6	9.9	597	100	AQ652626	AQ652626 Sheared D
C 6	35.2	9.8	511	23	AW659211	AW659211 96177 MAR
C 7	33.8	9.4	609	24	AW765044	AW765044 da87601.Y
C 8	33.4	9.3	341	5	AA696560	AA696560 GM07881.5
C 9	33.4	9.3	634	112	AQ950287	AQ950287 Sheared D
C 10	33.4	9.3	719	35	BE430706	BE430706 SUN007.CO
C 11	33	9.2	331	3	AA334604	AA334604 EST38844
C 12	33	9.2	489	6	AA859269	AA859269 UI-R-E0-C
C 13	33	9.2	713	25	AW955769	AW955769 EST367839
C 14	33	9.2	1101	121	CNS00KK2	AL077673 Drosophila
C 15	32.6	9.1	480	121	CNS00TNV	AL089929 Arabidops
C 16	32.6	9.1	741	36	BE534866	BE534866 601231410
C 17	32.2	8.9	426	36	C95499	C95499 C95499 Citr
C 18	32	8.9	380	23	AW599704	AW599704 ga91908.Y
C 19	32	8.9	400	14	AL361014	AL361014 AL361014
C 20	32	8.9	716	25	AW940039	AW940039 GH01388.3
C 21	32	8.9	1101	121	CNS00BXY	AL057978 Drosophila
C 22	31.8	8.8	421	21	AW248410	AW248410 2820429.5
C 23	31.6	8.8	440	3	AA298086	AA298086 EST113719
C 24	31.6	8.8	457	8	AL109978	AL109978 GH09394.5
C 25	31.6	8.8	625	6	AA736166	AA736166 HL08104.5
C 26	31.4	8.7	424	20	AW233549	AW233549 f37908.x
C 27	31.4	8.7	444	32	BE017212	BE017212 fk77h12.Y
C 28	31.4	8.7	551	34	BE201942	BE201942 f103009.x
C 29	31.4	8.7	606	34	BE201791	BE201791 fk98007.x
C 30	31.4	8.7	623	36	BE543934	BE543934 601070441
C 31	31.2	8.7	484	19	AV588611	AV588611 AV588611
C 32	31.2	8.7	509	36	BE497852	BE497852 WHE0957_E
C 33	31.2	8.7	555	113	AQ951351	AQ951351 Sheared D
C 34	31.2	8.7	730	121	CNS012MS	AL102118 Drosophila
C 35	31.2	8.7	1049	123	CNS04SBU	AL304995 Tetraodon
C 36	31.2	8.7	1052	122	CNS05BPK	AL236801 Tetraodon
C 37	31	8.6	538	15	AU066586	AU066586 AU066586
C 38	31	8.6	753	40	Z78358	Z78358 HS278358 Hu
C 39	31	8.6	960	34	BE283129	BE283129 601103594
C 40	31	8.6	1006	121	CNS005YL	AL066088 Drosophila
C 41	30.8	8.6	359	35	BE400123	BE400123 AWB011.D0
C 42	30.8	8.6	421	5	AA625956	AA625956 zu92e03.s
C 43	30.8	8.6	457	21	AW342909	AW342909 f180f02.Y
C 44	30.8	8.6	486	36	BE499631	BE499631 WHE0962_C
C 45	30.8	8.6	524	7	AA898051	AA898051 NCC2A117

ALIGNMENTS

RESULT 1
 BB155217/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

BB155217 276 bp mRNA EST 29-JUN-2000
 BB155217 RIKEN full-length enriched, 16 days neonate thymus Mus musculus cDNA clone A130024C22 3', mRNA sequence.
 BB155217
 BB155217.1 GI:8811147
 EST.
 house mouse
 Mus musculus

REFERENCE
 AUTHORS
 1 (bases 1 to 276)
 Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Itoh,M., Iwazawa,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Iizawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Tagawa,A., Takahashi,F., Tomimaga,N., Toya,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I., Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.

RIKEN MOUSE ESTS (Konno,H., et al.)
 Unpublished (2000)
 Contact: Yoshihide Hayashizaki
 Genome Exploration Research Group, Life Science Tsukuba Center,
 Genome Science Laboratory
 The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
 Tel: +81-298-36-9013
 Fax: +81-298-36-9098
 Email: genome-res@rtc.riken.go.jp/
 URL: http://genome.rtc.riken.go.jp/
 N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 Thermotabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
 Tomaru,Y., Carninci,P., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Y. and Hayashizaki,Y.
 Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
 Carninci,P. and Hayashizaki,Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
 Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

TITLE

JOURNAL

COMMENT

FEATURES

source

Location/Qualifiers
 1..276
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="A130024C22"
 /clone.lib="RIKEN full-length enriched, 16 days neonate thymus"
 /tissue_type="thymus"
 /dev_stage="16 days neonate"
 /lab_host="DH10B"
 /note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGCAGAGAGGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of

Fax: 402 762 4390

Email: smithemall.marc.usda.gov

Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -minmatch 12 options.

PCR Primers

FORWARD: AGGAACACGCTATGACCAT

BACKWARD: GTTTCACGTCAGCAGC

Plate: 83 row: F column: 22

Seq primer: ATTAGGTGACACATATAG.

Location/Qualifiers

1. 511

/organism="Bos taurus"

/db_xref="taxon:9913"

/clone_lib="MARC 1BOV"

/tissue_type="pooled"

/lab_host="DH10B"

/note="Vector: PCMV SPORF6; Site_1: XbaI; Site_2: XhoI;

Library made from pooled tissue from lymph node, ovary,

fat, hypothalamus, and pituitary."

124 a 119 c 184 g 84 t

BASE COUNT

ORIGIN

Query Match 9.8%; Score 35.2; DB 23; Length 511;
Best Local Similarity 50.6%; Pred. No. 1.8;
Matches 85; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 185 ctaaacggcgtgttaagcaagatgagcagaagcccaaacggaagcgtaacattggcgctg 244

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 95 CCAAGGAGGGTGTGTGGTGGCGGTGGAAGAACCAAGCAGGAGGAGTGACAGAGGCAGCTG 154

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 245 ttaaaccaacactgttgccgtccacgtaccacttccataaaacaggaagcactggagc 304

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 155 AGAAGACCAGGAGGTGTCTGTATGTGGGAGCTAAGACCAAGAGGGGTGTGTGCAGA 214

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 305 cagcgcaatcagtcacacacagtcgttgaggagagcgccgcttga 352

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 215 GTGTGACTTCAGTGGCTGAGAGACCAGGAGGAGCCCAAGCCGCTGA 262

RESULT 7
LOCUS AW765044

AW765044 609 bp mRNA EST 22-JUN-2000

da87e01.y1 Xenopus laevis tadpole stage 24 Xenopus laevis cDNA

clone IMAGE: 5' similar to gb:L11284 DUAL SPECIFICITY

MITOGEN-ACTIVATED PROTEIN KINASE 1 (HUMAN); gb:L02526 Mouse

protein kinase (MOUSE); gb:dbj|D13700.1|XELMAPKK Xenopus laevis

mRNA for MAP kinase, complete (XENOPUS); mRNA sequence.

ACCESSION AW765044

AW765044.1 GI:7697005

EST.

ORGANISM African clawed frog.

Xenopus laevis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;

Xenopodinae; Xenopus.

1 (bases 1 to 609)

Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.

, Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person

, B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,

Waterston,R. and Wilson,R.

WashU Xenopus EST project, 1999

Unpublished (1999)

Contact: Sandy Clifton, Ph.D.

WashU Xenopus EST project, 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Library constructed by B. Korn PhD. and S. Henze DNA Sequencing by:

Washington University Genome Sequencing Center

Clone distribution: Xenopus clones from this library are available through the RessourcenzentrumPrimarDatenbank, Berlin, Germany (web address: www.rzpd.de)

Seq primer: -40RP from Gibco

High quality sequence stop: 411.

FEATURES
source

1. 609

/organism="Xenopus laevis"

/db_xref="taxon:8355"

/clone="IMAGE:"

/clone_lib="Xenopus laevis tadpole stage 24"

/sex="mixed"

/tissue_type="whole tadpole"

/dev_stage="stage 24"

/lab_host="DH10B"

/note="Vector: PCMVSPORT6; Site_1: NotI; Site_2: SalI;

Method of cloning used: directed ligation; insertCheck:

95.8 blue and 4.2 white. Library constructed by Dr.

Bernhard Korn and Sabine Henze. Note: This is a Xenopus

Gene Collection (XGC) library."

175 a 143 c 164 g 127 t

BASE COUNT

ORIGIN

Query Match 9.4%; Score 33.8; DB 24; Length 609;
Best Local Similarity 52.5%; Pred. No. 5.2;
Matches 74; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 164 ccccgctcaagcgcgagcagcctaaaccgctgttaagcaagatgagcagaagcccaaac 223

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 60 CTCACCTCACTCCCAACATCGCTAAAAGAGAGCCCTACGCCCATACAGCTGAATCCCAACC 119

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 224 gcaagcgtcacattggcgtttaagcacaacagctgttgccgtccacgtaccacttccta 283

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 120 CCGAAGGAGTGTGTGAACGGGACCCCTACAGCCGAGACAACTTGAAGCTCTGCAGA 179

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 284 aaaaacaggagcactggagc 304

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 180 AAAAGTTGAAGAGCTTGAGC 200

RESULT 8
LOCUS AA696560/c

AA696560 341 bp mRNA EST 28-NOV-1998

GM07881.5prime GM Drosophila melanogaster ovary Bluescript

Drosophila melanogaster cDNA clone GM07881 5prime, mRNA sequence.

ACCESSION AA696560

AA696560.1 GI:2699180

EST.

ORGANISM fruit fly.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 341)

Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein

,P., Lewis,S. and Rubin,G.M.

BDGP/HHMI Drosophila EST Project

Unpublished (1997)

Contact: Harvey, D.

G. M. Rubin-Molecular and Cell Biology

University of California Berkeley

539 LSA, Berkeley, CA 94720-3200, USA

Fax: 510 643 9947

Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu

Plate: 78 row: G column: 9

High quality sequence stop: 156.

Location/Qualifiers

1. 341

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone="GM07881"

/clone_lib="GM Drosophila melanogaster ovary Bluescript"

```

/sex="female"
/dev_stage="newly enclosed females: germarium-stage 6"
/lab_host="SOLR"
/notes="Organ: ovary; Vector: Bluescript SK; Site_1: EcoRI;
Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
Synthesis kit. Oligo dt-primed and directionally cloned at
EcoRI and XhoI in Bluescript SK(-/-)"
BASE COUNT      73 a   82 c   97 g   89 t
ORIGIN

```

```

Query Match      9.3%; Score 33.4; DB 5; Length 341;
Best Local Similarity 55.7%; Pred. No. 6.2;
Matches 64; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 134 tcgtacactaccacgggtagtcggtatccccctcgaagcgagcagcctaaccgg 193
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 240 TGGCGGCTACCTCGGTCGGAACGACATCTCGCATAGCGCGCATGGCCAACTGG 181

```

```

QY 194 ctgttaagcaagtagcagcaagcccaacgcaagcgctcacattggcgtgttaa 248
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 180 CTGCTATGGAGGAGGAGCAGCAAGCAACTGGCAACGGAACATTGTGAATTAAA 126

```

```

RESULT 9
A0950287      634 bp   DNA      GSS      27-JAN-2000
LOCUS
DEFINITION   Sheared DNA-37L6.TF Sheared DNA Trypanosoma brucei genomic clone
ACCESSION   A0950287
VERSION     A0950287.1 GI:6773552
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei.
ORGANISM    Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.

```

```

REFERENCE 1 (bases 1 to 634)
AUTHORS   El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
            Gerrard,C., Leach,V., de Jong,P., Ullu,E., Melville,S., Donelson,J.,
            Fraser,C., and Adams,M.
TITLE     Determination of clone end sequences from Trypanosoma brucei GUTat
            10.1 sheared DNA library
JOURNAL   Unpublished (1999)
COMMENT   Other GSSs: Sheared DNA-37L6.TF
            Contact: Najib M. El-Sayed
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: nelsayed@tigr.org
            Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
            DNA library constructed at TIGR. Clones will be available for
            distribution through ATCC. Sheared DNA end sequences search page:
            http://www.tigr.org/tldb/mdb/tbdb/.
            Seq primer: M13-Reverse
            Class: Shotgun.

```

```

FEATURES
Source      Location/Qualifiers
1..634
/organism="Trypanosoma brucei"
/strain="TREU927/4 GUTat 10.1"
/db_xref="taxon:5691"
/clone_lib="Sheared DNA-37L6"
/clone="Sheared DNA"
/notes="Vector: pUC18; Site_1: SmaI; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
sheared to give a tight size distribution (approx 2 kb).
The v + i method used for the library construction is
described in detail in Smith, H.O. and Venter, J.C.
(Making small insert libraries for whole genome shotgun
sequencing projects. In Genome Sequencing: A Practical

```

```

Approach, eds. M. Vaudin and B. Barell, Oxford University
Press, 1999)."
BASE COUNT      203 a   122 c   170 g   138 t   1 others
ORIGIN

```

```

Query Match      9.3%; Score 33.4; DB 112; Length 634;
Best Local Similarity 58.6%; Pred. No. 6.9;
Matches 58; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

```

```

QY 250 ccaacagctgttgctccacgtaccactctctaaacacaggaagcactgagcagcg 309
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 348 CCAACAGTTTATCCTTCAGGGACCACTTCTTACACCACAGCAGCAGGAGGAATGG 407

```

```

QY 310 caatcagtcaccacaacagctgttgaggaggaagggccgcg 348
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 408 CTCTGTGATTCGCTGTGAGACGGTGAAGGAGGTGAAGCC 446

```

```

RESULT 10
BE430706      719 bp   mRNA      EST      24-JUL-2000
LOCUS
DEFINITION   SUN007.C09F991221 ITEC SUN Wheat cDNA Library Triticum aestivum
            cDNA clone SUN007.C09, mRNA sequence.
ACCESSION   BE430706
VERSION     BE430706.1 GI:9428661
KEYWORDS    EST.
SOURCE      bread wheat.
ORGANISM    Triticum aestivum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Triticum.

```

```

REFERENCE 1 (bases 1 to 719)
AUTHORS   Anderson,O.A., Appeals,R., Bailey,P., Blake,T., Close,T., Cloutier
            ,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,
            Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P.,
            Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y.,
            Pengridi,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M.,
            Sorrells,M., Warburton,M., and Wenzel,G.
TITLE     International Tritecae EST Cooperative (ITEC): Production of
            Expressed Sequence Tags for Species of the Triticeae
JOURNAL   Unpublished (2000)
COMMENT   Contact: Shariflou M
            Plant Breeding Institute, University of Sydney
            107 Cobbitty Rd., Cobbitty NSW 2570 Australia
            Tel: 61 2 9351 8803
            Fax: 61 2 9351 8851
            Email: msharif@pop.usyd.edu.au
            International Triticeae EST Cooperative (ITEC)
            http://wheat.pw.usda.gov/genome.

```

```

FEATURES
Source      Location/Qualifiers
1..719
/organism="Triticum aestivum"
/db_xref="taxon:4565"
/clone="SUN007.C09"
/clone_lib="ITEC SUN Wheat cDNA Library"
BASE COUNT      115 a   264 c   184 g   142 t   14 others
ORIGIN

```

```

Query Match      9.3%; Score 33.4; DB 35; Length 719;
Best Local Similarity 55.7%; Pred. No. 7.1;
Matches 64; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

```

```

QY 63 ctcaacttcgcaactccgacgtgtctgatttcgacgtatttcgagctgtttctaa 122
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 330 CTCACCTCCGCTTTCCGGCTTCTCCGACTCCGGCCGGGATTCTACAGGTCTACGG 389

```

```

QY 123 tgcattactgtctacataccaccgggtagtcgctatcccccgctcaagcgc 177
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 390 CGACGTATTCGACAAAGGTATACGCCAGGAGTGCCTTCATGCCCGCCGATGGC 444

```

```

RESULT 11

```



```

AA334604/C
LOCUS      AA334604      331 bp      mRNA
DEFINITION EST38844 Embryo, 9 week Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION  AA334604
VERSION     AA334604.1 GI:1986848
KEYWORDS   EST.
SOURCE      human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 331)
AUTHORS    Adams,N.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
            ,C.J., Lee,N.H., Kirkness,E.F., Weinstein,K.G., Gocayne,J.D., White
            ,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
            Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
            ,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
            Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M.,
            Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palancques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimker,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.,
            Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
            Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H., Raymond,L.,
            Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
            ,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
            Venter,J.C.
            Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
            Nature 377 (6547 Suppl), 3-174 (1995)
            96026280
            Other ESTs: THC174272
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
            Location/Qualifiers
                1..331
                /organism="Homo sapiens"
                /db_xref="ATCC (inhost):136288"
                /db_xref="taxon:9606"
                /clone_lib="Embryo, 9 week"
                /dev_stage="embryo, 9 wks"
                /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                XhoI"
BASE COUNT      89 a 72 c 81 g 89 t
ORIGIN
Query Match      9.2%; Score 33; DB 3; Length 331;
Best Local Similarity 51.0%; Pred. No. 8.2;
Matches 78; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
QY 153 agtgcgtatgccccgcgtcaagcgagcagcctaaacccggctgttaagcaagatgagca 212
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 156 AGTTCAGCAGACCCCTGTACAGCTGTTCCTCCGGCCAACTGGGATCCTTAAGGAACAGGCTG 97
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 213 gaagcccaaacgggaagcgctcaattggctgttaagccaacagctgttggcgctccagct 272
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 96 GAAGAGCCAGGGCGCTCCCTCAATAATGCCTAAATAAACCAACATATCTATGCCGAGGCC 37
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 273 accattctctaaaaaacagggaagcactggagcc 305
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 36 ATCCCTCTCTCAATGCCATGGTCCAGGGAGTCC 4
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
AA859269      489 bp      mRNA
DEFINITION EST
ACCESSION  AA859269.1 GI:4229623
VERSION     AA859269
KEYWORDS   EST.
SOURCE      Norway rat.
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE  1 (bases 1 to 489)
AUTHORS    Bonaldo,M.F., Lennon,G. and Soares,M.B.
            Normalization and subtraction: two approaches to facilitate gene
            discovery
            Genome Res. 6 (9), 791-806 (1996)
            97044477
            On Mar 10, 1998 this sequence version replaced gi:2948620.
            Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence tag present in the cDNA between the NotI site and the
            oligo-dT track served to identify it as a clone from the normalized
            adult 8-Day-Embryo library. cDNA Library Preparation: M. Fatima
            Bonaldo, Ph.D. Clone distribution: clones will be available through
            Research Genetics This clone is also available through the
            I.M.A.G.E. Consortium at LLNL (infoimage.llnl.gov). IMAGE
            ID=1777539 The following repetitive elements were found in this
            cDNA sequence: 1-25, >AT-rich#Low_complexity
            Seq primer: M13 Forward
            POLYA=No.
            Location/Qualifiers
                1..489
                /organism="Rattus norvegicus"
                /strain="Sprague-Dawley"
                /db_xref="taxon:10116"
                /clone="UI-R-E0-cj-g-01-0-UI"
                /clone_lib="UI-R-E0"
                /dev_stage="embryonic"
                /lab_host="DH10B (Life Technologies)"
                /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                polylinker; Site_1: NotI; Site_2: EcoRI; This library
                consists of a mixture of individually tagged normalized
                libraries constructed from 8, 12 and 18-day embryo. The
                tag is a string of 3-5 nucleotides present between the
                Not I site and the oligo-dT track which allows
                identification of the library of origin of a clone within
                the mixture."
BASE COUNT      87 a 108 c 138 g 156 t
ORIGIN
Query Match      9.2%; Score 33; DB 6; Length 489;
Best Local Similarity 48.2%; Pred. No. 8.8;
Matches 93; Conservative 0; Mismatches 100; Indels 0; Gaps 0;
QY 156 ggcgtatgccccgcgtcaagcgagcagcctaaacccggctgttaagcaagatgagcagaa 215
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 336 GCTACATGAATACAGACCCCGATACCTTATGCATGATGAGATGGCATGTGGCACC 277
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 216 gcccaaacgcaagcgctcaattggcggttaagccaacagctgttggcgctccagctacc 275
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 276 AGCCATACGCCAGCAGCCCAATGCAGACCCACCCACAGTACATGATGTACAGGCC 217
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 276 acttctaaaaaacagggaagcactggagcgagcgcaatcagtcaccacacagctgttgg 335
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 216 CCGACACCGGTACATGAACAGCGGATGTCCTCAACAGCTCTCAACGGGTCTTACA 157
QY 336 ggagaagccgcc 348
Db 156 TGAGAAGGTAGAC 144

RESULT 13
AW955769/c
LOCUS
DEFINITION EST 713 bp mRNA EST 01-JUN-2000
ACCESSION EST367839 MAGE resequences, MAGD Homo sapiens cDNA, mRNA sequence.
VERSION AW955769
KEYWORDS AW955769.1 GI:8145452
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspari,R., Gay,C., Holt
I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
Quackenbush,J.
TITLE Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
JOURNAL Unpublished (2000)
COMMENT Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 83
Seq primer: Reverse.
FEATURES
source Location/Qualifiers
1..713
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGD"
/notes="Vector: pBluescriptSKm"
BASE COUNT 196 a 133 c 167 g 217 t
ORIGIN

Query Match 9.2%; Score 33; DB 25; Length 713;
Best Local Similarity 51.0%; Pred. No. 9.4;
Matches 78; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 153 agtgcgctatgccccgcgtaagcgcgagcagcctaaaccgctgttaagcaagatgagca 212
Db 155 AGTTCAGCAGACCCCTTGTACAGCTGTTCGCGCCAACTGGATCCTTAAGGAACAGCGTG 96
QY 213 gaagcccaaacgcgacgctcacattgggtgttaagcccaacagctgttggcgtccacgt 272
Db 95 GAAGAGCGAGGGCGCTCCCTCACAATAATGCTTAAATACCAACATATCTATGCCACAGCC 36
QY 273 accattcctataaaacagagaagcactggagcc 305
Db 35 ATCCCTCTCTCAATGCCATGTGTCAGGGAGTC 3

RESULT 14
CNS00KK2 1101 bp DNA GSS 03-JUN-1999
LOCUS
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC:
BACr17F22 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL077673
VERSION AL077673.1 GI:4957249
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

```

```

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1..(bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammosier in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
source Location/Qualifiers
1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACr17F22"
/notes="end : T7"
BASE COUNT 191 a 94 c 163 g 151 t 502 others
ORIGIN

Query Match 9.2%; Score 33; DB 121; Length 1101;
Best Local Similarity 8.8%; Pred. No. 10;
Matches 27; Conservative 144; Mismatches 136; Indels 0; Gaps 0;

QY 15 cagcgcgctgagagggtgttaagaaggtgttgatgaggcaagcaacctcaactccgc 74
Db 795 MNGMMMMVMKMGMMGGMGKGMGKGMKGMKGMKGMKGMKGMKGMKGMKGMKGMKGMK 854
QY 75 aactccgacgtgtctgatttcgacgtctgatttcgacgtcttcttaactgcttctgt 134
Db 855 MNNDKNNNNCKTKOMNNKKYMMMMMMMAKMMHKKMMKMMKMMKMMKMMKMMKMMKMM 914
QY 135 cgtacactaccacccgggtagtcgctgtagtccccgcgtaagcgcgagcagcctaaaccgc 194
Db 915 CKKKMAKTCBMMMAKMAKVGSMAKDKMDGMAGMAMMNGCMGCMGMMAGCKMKMMMMG 974
QY 195 tgttaagcaagatgagcagaagcccaaacgcgacgctcacattggctgtttaagccaac 254
Db 975 AGVAMGGTAGKMMKMGTMAMMAKMAKMAKMAKMAKMAKMAKMAKMAKMAKMAKMAKMA 1034
QY 255 agctgttggtccacgtaccacttctctaaacaaagcagcactggagcagcgaatc 314
Db 1035 MKMKHVMHMMKMMMDMDHANNKMKACMMKMKMMKMMKMMKMMKMMKMMKMMKMMKMM 1094
QY 315 agtccca 321
Db 1095 MADKMM 1101

RESULT 15
CNS00TNV/c 480 bp DNA GSS 28-JUN-1999
LOCUS
DEFINITION Arabidopsis thaliana genome survey sequence T7 end of BAC r101 of
TAMU library from strain Columbia of Arabidopsis thaliana, genomic
survey sequence.
ACCESSION AL089929
VERSION AL089929.1 GI:5291069
KEYWORDS GSS.
SOURCE thale cress.

```

Search completed: December 18, 2000, 15:12:46
Job time: 18733 sec